

STIC-Biotech/ChemLib

91690

From: Slobodyansky, Elizabeth  
Sent: Wednesday, April 16, 2003 2:30 PM  
To: STIC-Biotech/ChemLib  
Subject: 10/067,291

Please search for case 10/067,291:

SEQ ID NOs: 1 and 2 against commercial and interference databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652  
CM1 10D11  
703-306-3222

mail box 10D01

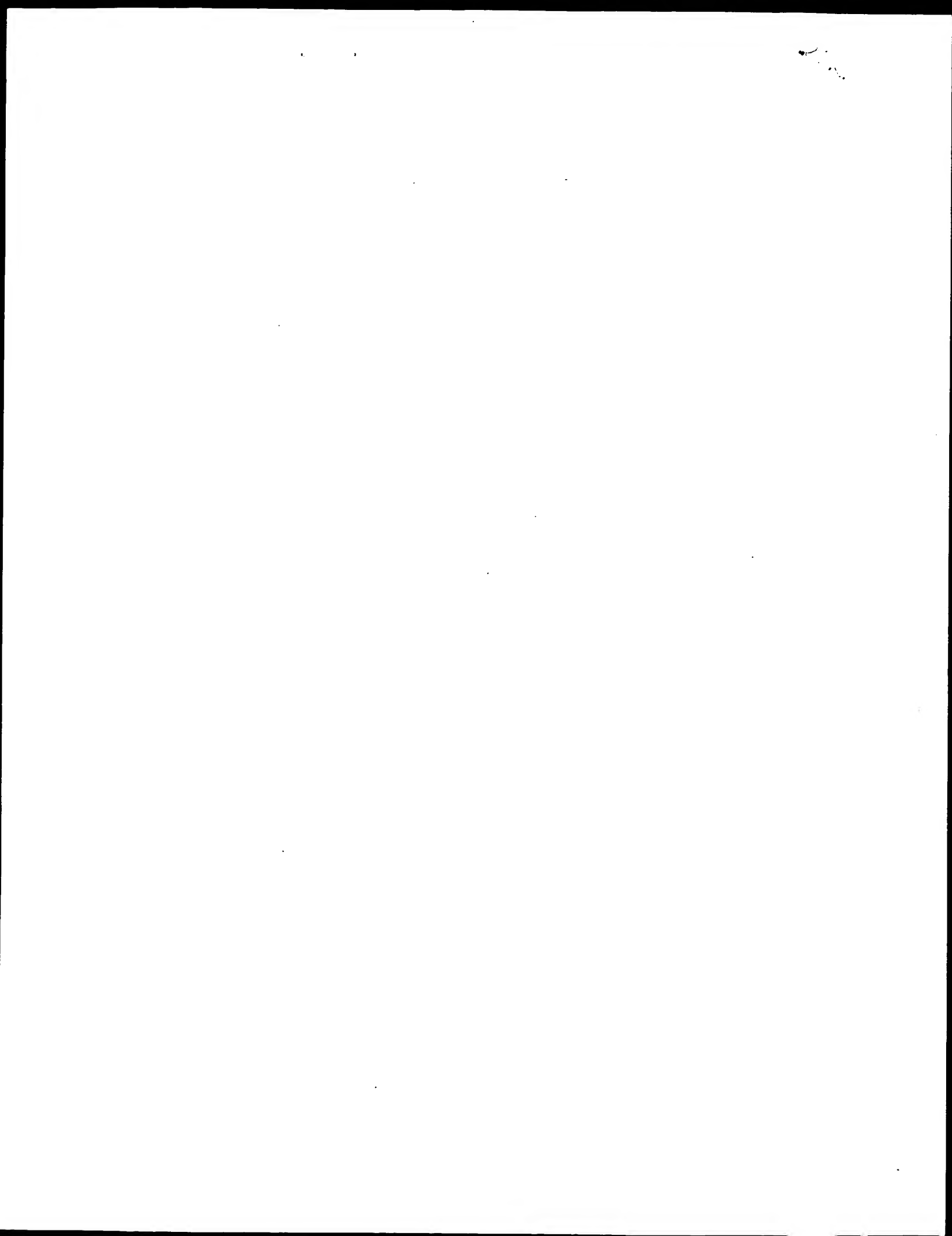
CRFE

POINT OF CONTACT:  
PAUL SCHULWITZ  
TECHNICAL INFO. SPECIALIST  
CM1 6B06 TEL. (703) 305-1954

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 4/17  
Date Completed: 4/25  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: 1  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_





XX Claim 2; Page 30-32; 54pp; English.  
 PS The present sequence is the amino acid sequence of a protein from  
 XX Mycobacterium aurum which acts as a biocatalyst. It is able to catalyze  
 CC transamination stereoselectively to produce an optically active amino  
 CC compound.  
 CC  
 SQ Sequence 339 AA;  
 Query Match 100.0%; Score 1764; DB 21; Length 339;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-168;  
 Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 MTALSDLGTSNLAVERPGAIRREDTPAGSVIYQSDVELDTSSPPAGGVAMIEGTYLPAREA 60  
 DB 1 MTALSDLGTSNLAVERPGAIRREDTPAGSVIYQSDVELDTSSPPAGGVAMIEGTYLPAREA 60  
 OY 61 KISIFDTGFGHSDLTYYVAHVHGNIFRLGDHDLRLDGLDASKRLDAGYSKDELAETTK 120  
 DB 61 KISIFDTGFGHSDLTYYVAHVHGNIFRLGDHDLRLDGLDASKRLDAGYSKDELAETTK 120  
 OY 121 CYSMSQRLRESFVNLVTYRGYGRKRGKDLKTHQYIYAIPYLMAPPAPQIFGTATY 180  
 DB 121 CYSMSQRLRESFVNLVTYRGYGRKRGKDLKTHQYIYAIPYLMAPPAPQIFGTATY 180  
 OY 181 PRHVRARGRTVPTIKNTYQMGDLTAASFEAKDRCARFALILDSDNCVAEGGFNVCIVK 240  
 DB 181 PRHVRARGRTVPTIKNTYQMGDLTAASFEAKDRCARFALILDSDNCVAEGGFNVCIVK 240  
 OY 241 DGLKASPSRNALPGITRTKTYFELADQMGIEATLRDVTSRRLYDADELMATVTTAGVTPIN 300  
 DB 241 DGLKASPSRNALPGITRTKTYFELADQMGIEATLRDVTSRRLYDADELMATVTTAGVTPIN 300  
 OY 301 SLDGEAVNGEGPPLTVAIRDRFALMDPEGPPIETIEY 339  
 DB 301 SLDGEAVNGEGPPLTVAIRDRFALMDPEGPPIETIEY 339  
 RESULT 2  
 AAB26104  
 ID AAB26104 standard; Protein: 339 AA.  
 AC AAB26104;  
 XX 15-JAN-2001 (first entry)  
 DE Mycobacterium aurum stereoselective transamination catalyst mutant.  
 KW Transamination; optically active amino compound; isomer; mutant; muten.  
 XX  
 OS Mycobacterium aurum.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 2 /note= "wild-type thr substituted by ala"  
 FT  
 XX EPI038953-A1.  
 XX  
 PD 27-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000EP-0400701.  
 XX  
 PR 19-MAR-1999; 99JP-0075511.  
 PR 30-MAR-1999; 99JP-0088634.  
 XX  
 PA (SUMO ) SUMITOMO CHEM CO LTD.  
 XX  
 PI Takashima Y, Mitsuda S, Wiser M;  
 XX  
 DR WPI: 2000-629576/61.  
 DR N-PSDB; AAA94040.

XX Novel Mycobacterium protein capable of catalyzing transamination stereo  
 PT selectively; useful for producing optically active amino compounds,  
 PT comprises a sequence of 339 amino acids  
 PS Claim 3; Page -: 54pp; English.  
 XX  
 XX The present sequence is a mutant version of the amino acid sequence of a  
 CC protein from Mycobacterium aurum which acts as a biocatalyst. It is able  
 CC to catalyze transamination stereoselectively to produce an optically  
 CC active amino compound.  
 CC Note: This sequence is not shown in the specification but is derived  
 CC from the Mycobacterium aurum shown in SEQ ID NO: 1 (see AAB26100).  
 CC  
 SQ Sequence 339 AA;  
 Query Match 99.7%; Score 1759; DB 21; Length 339;  
 Best Local Similarity 99.7%; Pred. No. 5.5e-168;  
 Matches 338; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 MTALSDLGTSNLAVERPGAIRREDTPAGSVIYQSDVELDTSSPPAGGVAMIEGTYLPAREA 60  
 DB 1 MAALSDLGTSNLAVERPGAIRREDTPAGSVIYQSDVELDTSSPPAGGVAMIEGTYLPAREA 60  
 OY 61 KISIFDTGFGHSDLTYYVAHVHGNIFRLGDHDLRLDGLDASKRLDAGYSKDELAETTK 120  
 DB 61 KISIFDTGFGHSDLTYYVAHVHGNIFRLGDHDLRLDGLDASKRLDAGYSKDELAETTK 120  
 OY 121 CYSMSQRLRESFVNLVTYRGYGRKRGKDLKTHQYIYAIPYLMAPPAPQIFGTATY 180  
 DB 121 CYSMSQRLRESFVNLVTYRGYGRKRGKDLKTHQYIYAIPYLMAPPAPQIFGTATY 180  
 OY 181 PRHVRARGRTVPTIKNTYQMGDLTAASFEAKDRCARFALILDSDNCVAEGGFNVCIVK 240  
 DB 181 PRHVRARGRTVPTIKNTYQMGDLTAASFEAKDRCARFALILDSDNCVAEGGFNVCIVK 240  
 OY 241 DGLKASPSRNALPGITRTKTYFELADQMGIEATLRDVTSRRLYDADELMATVTTAGVTPIN 300  
 DB 241 DGLKASPSRNALPGITRTKTYFELADQMGIEATLRDVTSRRLYDADELMATVTTAGVTPIN 300  
 OY 301 SLDGEAVNGEGPPLTVAIRDRFALMDPEGPPIETIEY 339  
 DB 301 SLDGEAVNGEGPPLTVAIRDRFALMDPEGPPIETIEY 339  
 RESULT 3  
 AAB82241  
 ID AAB82241 standard; Protein: 325 AA.  
 AC AAB82241;  
 XX 26-FEB-1999 (first entry)  
 DE A protein with stereoselective transaminase activity.  
 KW Stereoselective; transaminase activity; optically active amine;  
 KW R-chirality; ketone; synthetic intermediate; pharmaceutical;  
 XX agrochemical.  
 XX  
 OS Arthrobacter sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1 /note= "optional"  
 FT  
 XX MO9848030-A1.  
 XX  
 PD 29-OCT-1998.  
 XX  
 PF 20-APR-1998; 98WO-JP01814.  
 PF  
 PR 23-APR-1997; 97JP-0121732.  
 PR  
 XX



PA (KANE) KANEKA CORP.  
 XX Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;  
 PI Ogura M, Yamada Y;  
 XX WPI, 1998-583664/49.  
 DR N-PSDB; AAV73146.  
 XX  
 PT Production of optically active amine(s) from ketone(s) and an amino  
 PT donor - using a transaminase derived from an Arthrobacter species.  
 XX  
 PS Claim 22; page 56-57; 84pp; Japanese.  
 XX  
 CC The present sequence represents a protein with stereoselective  
 CC transaminase activity. The transaminase is used to produce  
 CC the optically active amines (having R-chirality) of the  
 CC invention. A ketone is treated with a transaminase in the  
 CC presence of a primary amine as amino donor to produce the  
 CC amines. The chiral amines are synthetic intermediates for  
 CC use as pharmaceuticals and agrochemicals.  
 CC  
 XX Sequence 325 AA;

Query Match 48.0%; Score 847; DB 19; Length 325;  
 Best Local Similarity 51.0%; Pred. No. 2.2e-76;  
 Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

OY 30 IYSDYELDTSSPPAGVAMIEGEYLPAEAKISIFDTGFGHSDLTYYVAHWGNIPL 89  
 ||||| :  
 DB 16 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDGIHSDVITYVFWNGNAFRL 75  
 :  
 OY 90 GDHLDRLDAGSKRLDAGYKDELAETTKCVSMQSOLRESFVNLVTRGYGRKGEKL 149  
 :  
 DB 76 DDHIERLFSAESMRILPPLTQDEVKEIAELVAKTELREAFVSITRGYSTPGERDI 135  
 :  
 OY 150 SKLTHOYIYAIPLMAFPFAEQIFGTATVPRHVRBAGRTVPTIKNYOMGLTASAF 209  
 :  
 DB 136 TKHRPQYMYAVPYQWIVPEFDRIRDGVAHMAVQSVRRTPRSSIDPOVKNFQMGDLIRAVQ 195  
 :  
 OY 210 EAKDGRATAILDSDNCVAGPFGFNVCIYDGLASPSRNALPGITRKTYFELADOMGI 269  
 ||||| :  
 DB 196 ETHDRGEAPLILDDGDLAEGSGFNVVYIKDGVVRSPGRALPGITRKTYLEIAESLGH 255  
 :  
 OY 270 EATLRDVTSRRLYDADELMATVAGTVPINSIDGAVNGEPGLVVALRDRFWALMDE 329  
 ||||| :  
 DB 256 EAILADITLAEILDADDEVLGCTTAGVWPVSVSDGNPISDVGPGITQSIIRRYWELANE 315  
 :  
 OY 330 PGPILETIEY 339  
 :  
 DB 316 SSSLITPVOY 325

RESULT 4  
 AAM87563  
 ID AAM87563 standard; Protein: 329 AA.

XX AAM87563;  
 AC 26-FEB-1999 (first entry)  
 DT  
 XX  
 DE A protein with stereoselective transaminase activity.  
 XX  
 DE Stereoselective; transaminase activity; optically active amine;  
 KM R-chirality; ketone; synthetic intermediate; pharmaceutical;  
 XX agrochemical.  
 XX Arthrobacter sp.  
 OS  
 PN W09848030-AI.  
 XX  
 PD 29-OCT-1998.  
 XX  
 PF 20-APR-1998; 98WO-JP01814.

XX 23-APR-1997; 97JP-0121732.  
 PR  
 XX (KANE) KANEKA CORP.  
 PA  
 XX Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;  
 PI Ogura M, Yamada Y;  
 XX WPI, 1998-583664/49.  
 DR N-PSDB; AAV83675.  
 XX  
 PT Production of optically active amine(s) from ketone(s) and an amino  
 PT donor - using a transaminase derived from an Arthrobacter species.  
 XX  
 PS Claim 22; page 56-57; 84pp; Japanese.  
 XX  
 CC The present sequence represents a protein with stereoselective  
 CC transaminase activity. The transaminase is used to produce  
 CC the optically active amines (having R-chirality) of the  
 CC invention. A ketone is treated with a transaminase in the  
 CC presence of a primary amine as amino donor to produce the  
 CC amines. The chiral amines are synthetic intermediates for  
 CC use as pharmaceuticals and agrochemicals.  
 CC  
 XX Sequence 329 AA;

Query Match 48.0%; Score 847; DB 19; Length 329;  
 Best Local Similarity 51.0%; Pred. No. 2.2e-76;  
 Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

OY 30 IYSDYELDTSSPPAGVAMIEGEYLPAEAKISIFDTGFGHSDLTYYVAHWGNIPL 89  
 ||||| :  
 DB 20 ITYSDYELDPANPLAGGAAWIEGAFVPPSEARISIFDGIHSDVITYVFWNGNAFRL 79  
 :  
 OY 90 GDHLDRLDAGSKRLDAGYKDELAETTKCVSMQSOLRESFVNLVTRGYGRKGEKL 149  
 :  
 DB 80 DDHIERLFSAESMRILPPLTQDEVKEIAELVAKTELREAFVSITRGYSTPGERDI 139  
 :  
 OY 150 SKLTHOYIYAIPLMAFPFAEQIFGTATVPRHVRBAGRTVPTIKNYOMGLTASAF 209  
 :  
 DB 140 TKHRPQYMYAVPYQWIVPEFDRIRDGVAHMAVQSVRRTPRSSIDPOVKNFQMGDLIRAVQ 199  
 :  
 OY 210 EAKDGRATAILDSDNCVAGPFGFNVCIYDGLASPSRNALPGITRKTYFELADOMGI 269  
 ||||| :  
 DB 200 ETHDRGEAPLILDDGDLAEGSGFNVVYIKDGVVRSPGRALPGITRKTYLEIAESLGH 259  
 :  
 OY 270 EATLRDVTSRRLYDADELMATVAGTVPINSIDGAVNGEPGLVVALRDRFWALMDE 329  
 ||||| :  
 DB 260 EAILADITLAEILDADDEVLGCTTAGVWPVSVSDGNPISDVGPGITQSIIRRYWELANE 319  
 :  
 OY 330 PGPILETIEY 339  
 :  
 DB 320 SSSLITPVOY 329

RESULT 5  
 AAM87564  
 ID AAM87564 standard; Protein: 330 AA.

XX AAM87564;  
 AC 26-FEB-1999 (first entry)  
 DT  
 XX  
 DE A protein with stereoselective transaminase activity.  
 XX  
 DE Stereoselective; transaminase activity; optically active amine;  
 KM R-chirality; ketone; synthetic intermediate; pharmaceutical;  
 XX agrochemical.  
 XX Arthrobacter sp.  
 OS  
 PN W09848030-AI.  
 XX

PD 29-OCT-1998.  
 XX  
 PF 20-APR-1998; 98WO-JP01814.  
 XX  
 PR 23-APR-1997; 97JP-0121732.  
 XX  
 PA (KANE) KANEKA CORP.  
 XX  
 PI Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;  
 PI Ogura M, Yamada Y;  
 XX  
 DR WPI: 1998-583664/49.  
 DR N-PSDB; AAV83676.  
 XX  
 PT Production of optically active amine(s) from ketone(s) and an amino  
 PT donor - using a transaminase derived from an Arthrobacter species.  
 XX  
 PS Claim 22; Page 56-57; 84pp; Japanese.  
 XX  
 CC The present sequence represents a protein with stereoselective  
 CC transaminase activity. The transaminase is used to produce  
 CC the optically active amines (having R-chirality) of the  
 CC invention. A ketone is treated with a transaminase in the  
 CC presence of a primary amine as amino donor to produce the  
 CC amines. The chiral amines are synthetic intermediates for  
 CC use as pharmaceuticals and agrochemicals.  
 XX  
 SQ Sequence 330 AA;  
 Query Match 48.0%; Score 847; DB 19; Length 330;  
 Best Local Similarity 51.0%; Pred. No. 2.2e-76;  
 Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;  
 Oy 30 IQSDYELDTSSPFAGVAMIEGELTPEAKISIFDTGFGHSDLTVAHVHGNIFRL 89  
 Db 21 IYSOYELDPANPLAGAMIEGAFVPPSEARISIFDGYLHSDVTVFVHVMGNMFR 80  
 Oy 90 GDHLRLDGASKRLDAGYKDELAEITTKCVSKQRESFVNLVTGTYGKRGKEXDL 149  
 Db 81 DDIHRLSNESMRITPLQDEYKELTALVAKTELEAFVSVITRGYSSPGERDI 140  
 Oy 150 SKLFGVYIYAIPLYMAFPPEAQIFGTATVPRHVRAGRTVPTIKNYMGDLTAASF 209  
 Db 141 TKHROVYIMVPRYQWYVPRIRIDGVANVAOVSRRPSSIDPOYKNFQMGDLIRAVQ 200  
 Oy 210 EAKDGFARTAILDSNCVAGPFPNCIVDGLASPSRNALPGITRKTVFELADOMGI 269  
 Db 201 ETHDRGFAPRLIDGDLAEGSGFNVVYIKDGVVRSRPAALPGITRKTVLEIAESIGH 260  
 Oy 270 EATLDVYSRELXADDELMAVTAGCTVPINSIDGAVNGEPPPLVATIDRFWALMDE 329  
 Db 261 EATLDVYLAELDLDEVLGCTAGGVWPFVSDGNPISDGPPIQSIIRRWELNVE 320  
 Oy 330 PGLLETLEY 339  
 Db 321 SSSLLTFVQY 330  
 RESULT 6  
 AAY28430  
 ID AAY28430 standard; Protein; 310 AA.  
 XX  
 AC AAY28430;  
 XX  
 DT 15-FEB-2000 (first entry)  
 XX  
 DE Corn branched chain amino acid transaminase amino acid sequence.  
 XX  
 KW Corn; soybean; wheat; rice; dihydroxyacid dehydratase; leuc; leud;  
 KW branched chain amino acid transferase; biosynthetic enzyme; antibody;  
 KW 3-isopropylmalate dehydratase.  
 XX  
 OS Zee mays.

XX  
 PN WO9921880-A2.  
 XX  
 PD 06-MAY-1999.  
 XX  
 PF 20-OCT-1998; 98WO-US22081.  
 XX  
 PR 28-OCT-1997; 97US-0063423.  
 XX  
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.  
 XX  
 PI Falco SC, Cahoon RE, Hiltz WD, Kinney AJ, Rafalski JA;  
 PI WPI: 2000-022904/02.  
 DR N-PSDB; AAX89455.  
 XX  
 PT Nucleic acid fragments encoding branched chain amino acid biosynthetic  
 PT enzymes  
 XX  
 PS Claim 6; Fig 5; 102pp; English.  
 XX  
 CC AAY28418-Y28431 are fragments of corn, soybean, wheat and rice branched  
 CC chain amino acid transaminase, amino acid sequences. Sequences  
 CC AAX89442-X89465 are nucleic acid fragments that encode all or a  
 CC substantial portion of dihydroxyacid dehydratase, a branched chain amino  
 CC acid transferase, a leuc or a leud subunit of 3-isopropylmalate  
 CC dehydratase from wheat, corn, soybean or rice. These enzymes are involved  
 CC in biosynthesis and utilization of branched-chain amino acids. The  
 CC nucleic acid sequences can be used to alter the level of expression of a  
 CC branched chain amino acid biosynthetic enzyme in a host cell. They can  
 CC also be used to obtain a nucleic acid fragment encoding a branched chain  
 CC amino acid biosynthetic enzyme, and to identify inhibitors of a branched  
 CC chain amino acid biosynthetic enzyme. The encoded proteins may be used  
 CC to prepare antibodies for detecting the proteins in situ in cells, or in  
 CC vitro in cell extracts.  
 XX  
 SQ Sequence 310 AA;  
 Query Match 18.3%; Score 322.5; DB 21; Length 310;  
 Best Local Similarity 29.0%; Pred. No. 1e-23;  
 Matches 82; Conservative 54; Mismatches 134; Indels 13; Gaps 5;  
 Oy 47 VANIEGELYPAEBAKISIFDTGFGHSDLTVAHVHGNIFRLGDLRLDGLASKRLRD 106  
 Db 18 LVWVGDELLPRNSAKSVSDSVQGDVAWEGRLTIDGVFKLDEHDLDFSAKAMAFS 77  
 Oy 107 AGYSKDELAE-ITKCVSKQRESFVNLVTGTYGKRGKEXDLSTHQVYIYAIPLY- 164  
 Db 78 NVETRMKIDALIKTLIANGMFNNHILRLTR-----GKRVTSQMSPAFNLCCALIV 131  
 Oy 165 ---WAEPPAEQIFGTATVPRHVRAGRTVPTIKNYMGDLTAASFPAKDRGARTAIL 221  
 Db 132 LAEMKRPVYDNSHG-ITKYATATRRSPNSIDPKIHNNLNNILAKIEGNLAQEDNAE 190  
 Oy 222 LQSDNCVAGPFPNCIVDGLASPSRN-ALPGITRKTVFELADOMGTEATLRDVTSE 280  
 Db 191 LDRDGFVSERNTNINIMVKKGIYLPHADYCLPGITRATVMDLVKENVLHERRISSE 250  
 Oy 281 LVDADELMVTTAGCTVPINSIDGAVNGEPPPLVATIDRF 323  
 Db 251 FRADEVMTTGTGELTPVMIDGREGIDGKIGPVTQIKAV 293  
 RESULT 7  
 ABB92207  
 ID ABB92207 standard; Protein; 490 AA.  
 XX  
 AC ABB92207;  
 XX  
 DT 31-MAY-2002 (first entry)  
 XX  
 DE Herbicidally active polypeptide SEQ ID NO 1418.

KM Herbicidal; plant; agriculture; herbicide.  
 XX Arabidopsis thaliana.  
 OS  
 XX W0200210210-A2.  
 PN  
 XX 07-FEB-2002.  
 PD  
 XX 28-AUG-2001; 2001WO-EP09892.  
 PF  
 XX 28-AUG-2001; 2001WO-EP09892.  
 PR  
 XX 28-AUG-2001; 2001WO-EP09892.  
 PA (FARB) BAYER AG.  
 XX  
 XX Tietjen K, Weidner M;  
 PI  
 XX WPI; 2002-269010/31.  
 DR  
 XX  
 XX  
 PT Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms -  
 PS  
 XX Claim 5; SEQ ID NO 1418; 261pp + sequence listing; English.  
 CC  
 CC The invention relates to identifying target proteins  
 CC (AB890790-AB894016) for herbicidally active compounds, comprising  
 CC aligning and comparing nucleic acid or amino acid sequences from plant  
 CC with nucleic acid or amino acid sequences from non-plant organisms using  
 CC suitable search parameters, where plant sequences having an E-value  
 CC greater by a factor of 3 than the E-value of most similar non-plant  
 CC sequences are selected. The polypeptides or nucleic acids encoding them  
 CC are useful for identifying modulators. The identified modulators are  
 CC useful as herbicides.  
 CC  
 CC  
 SQ Sequence 490 AA;  
 Query Match 18.1%; Score 318.5; DB 23; Length 490;  
 Best Local Similarity 28.0%; Pred. No. 5.1e-23;  
 Matches 82; Conservative 51; Mismatches 123; Indels 37; Gaps 3;  
 QY 47 VAMIEGEVLPAAEAKISIFDTGFGHSDLTFTVAHVHGNIFRGHIDLRLDGASKRLDL 106  
 DB 232 LAMVGDEILPRAKAKSVSDSVYOGSDSWBGLRIKGIKIFLEHIDLRLDLSAKALAND 291  
 QY 107 AGYSKDELAETKCKVSMQLSRESEVNLVTTRGCGKRGKDKLSLTLQVYIAYIPIYMA 166  
 DB 292 NVPAEEVEKAEIPLRLITLVLAE-----WK 316  
 QY 167 FPPAEQIFGTATVPRHVRAGNRYDPTIKNYQMGDLTAASFQAKDRGARTAILLSDSN 226  
 DB 317 -PPVYDNEGIVLVYVTRRNSPNNLDSKIHNNMLNLILAKIESNNNAADAIMLDKG 375  
 QY 227 CVAEGGFVNCIVKDGKLSAPSRN-ALPGITRKYVEFLADQMGIENTLADVTSRELYDAD 285  
 DB 376 YVSETNATINEMVKKCVLPHPADYCLPQITRATVMEVLVYKENTILEERKISLSEHRTAN 435  
 QY 286 ELMAVTTAGGVTPINSIDGSAVNGEGRPLVAIRDRFALMDEPGPLIETIE 338  
 DB 436 EVVTTGTGMEISPVYKIDGRVIGDKGVGVTRTLQNAKKLTEDSGVPIPTYO 488  
 RESULT 8  
 ID AAW24255 standard; Protein; 303 AA.  
 AC AAW24255;  
 XX  
 XX 18-FEB-1998 (first entry)  
 DT  
 XX Pyrobaculum aerophilum branched chain aminotransferase.  
 DE  
 XX Branched chain aminotransferase; IM2/BCA; chiral compound.  
 KM

XX  
 OS Pyrobaculum aerophilum strain IM2.  
 XX  
 XX W09729187-A1.  
 PN  
 XX 14-AUG-1997.  
 PD  
 XX 21-JAN-1997; 97WO-US01094.  
 PF  
 XX 08-MAY-1996; 96US-0646590.  
 PR  
 XX 09-FEB-1996; 96US-0599171.  
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 XX  
 XX Swanson RV, Warren PV;  
 PI  
 XX WPI; 1997-415343/38.  
 DR  
 XX N-PSDB; AAT78779.  
 DR  
 XX  
 XX  
 PT New transaminase(s) and aminotransferase(s) derived from host cells  
 PT - used for producing enzymes, and hybridisation probes for a cDNA or  
 PT genomic library  
 PS  
 XX Claim 19; Fig 8; 95pp; English.  
 CC  
 CC This protein sequence comprises branched chain aminotransferase  
 CC IM2/BCA of Pyrobaculum aerophilum IM2, a thermophilic sulphur  
 CC bacteria that grows optimally at 100 deg C and pH 7.0 in low salt  
 CC medium. The IM2/BCA amino acid sequence was deduced from genomic  
 CC DNA (see AAT78779). Claimed thermostable transaminases and  
 CC aminotransferases (AAW24248-57) can be produced from native or  
 CC recombinant host cells for use with L- and/or D-amino acids for  
 CC production of optically pure chiral compounds used in the  
 CC pharmaceutical, agricultural and other industries. A method is  
 CC claimed for transferring an amino group from an amino acid to an  
 CC alpha-keto acid using a claimed enzyme.  
 CC  
 CC  
 SQ Sequence 303 AA;  
 Query Match 17.9%; Score 315; DB 18; Length 303;  
 Best Local Similarity 28.0%; Pred. No. 5.6e-23;  
 Matches 85; Conservative 62; Mismatches 113; Indels 44; Gaps 11;  
 QY 42 PPAAGVAMIEGEVLPAAEAKISIFDTGFGHSDLTFTVAHVHGN--IFRGHIDLRL 97  
 DB 3 PYAKYI-WLDRILKWEPAKIHVLTALHYGTSIFEGIRGYWGDNLVFLRLEHIDRMV 61  
 QY 98 DGASKRLDAGYSKDELAETKCKVSMQLSRE-----SFV-RLTYTRGCGKRGKDKL 149  
 DB 62 RSAKILGINITYTRREVQAIVLETIKANNFREDVYIRPVAFVASQIVT-----LDI 112  
 QY 150 SKLTHQVYIAYIPIYMAFPPEAQIFGTATVPRHVR-----RAGRNVDPPTIKNYQ--W 201  
 DB 113 RLNEVSLAVIYEP-----FG-KYLSPPNGIKATIVSWRRVHTMTLPVMAKIGIY 160  
 QY 202 GDLTAASFEADRGAKRTAILLSDSNCAVAGGFVNCIVKDGKLSAP--SRNALGGITRKT 259  
 DB 161 VNSVLAIVLEARSRGDELDLVNGVYVGESEGFIFVYRGGRLETPPVHEISLIGITRDT 220  
 QY 260 VEFLADQMGIENTLADVTSRELYDADELMAYTTAGVTPINSIDGSAVNGEGRPLVAI 319  
 DB 221 VIKISGDVGLNVEKPTIREVYTADEVLYVGTAAEITPVEVVDGRITIGKPGPIITTKI 280  
 QY 320 RDRF 323  
 DB 281 AELY 284  
 RESULT 9  
 ID AAW73375 standard; Protein; 282 AA.  
 AC AAW73375;  
 XX  
 XX  
 KM

XX 12-FEB-1999 (first entry)  
 DT S. aureus D-alanine transferase protein sequence.  
 XX  
 DE D-alanine transferase; gene therapy; immune response; respiratory tract;  
 KW bacterial infection; cardiac disease; gastrointestinal disorder;  
 KW central nervous system disorder; eye disease; urinary tract disorder.  
 XX Staphylococcus aureus.  
 OS  
 XX EP881285-A1.  
 XX 02-DEC-1998.  
 PD  
 XX 26-MAY-1998; 98EP-0304139.  
 XX  
 PR 29-MAY-1997; 97US-0865238.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Warren RL;  
 DR WPI: 1999-001391/01.  
 DR N-PSDB: AAV08568.  
 XX  
 PT New d-alanine transferase (dat) polypeptide and polynucleotide -  
 PT useful as diagnostic reagents and for prevention and treatment of  
 PT gastrointestinal, respiratory and CNS bacterial infections  
 XX  
 PS Claim 11; Page 26-27; 31pp; English.  
 XX  
 CC This sequence is the d-alanine transferase (dat) protein of the  
 CC invention. Dat polypeptides and polynucleotides are useful for diagnosing  
 CC susceptibility to diseases by detecting mutations or polymorphisms of the  
 CC dat gene. They can diagnose diseases caused by organisms comprising the  
 CC dat gene by detection at the nucleic acid level, and analysing for the  
 CC presence or amount of the dat polypeptide in tissue samples. Dat  
 CC polypeptides and polynucleotides can be used to screen for agonists and  
 CC antagonists which can be used in treatment to enhance or block dat  
 CC activity. Gene therapy may also be used to affect endogenous dat  
 CC polypeptide production. An immunological response can be induced by  
 CC administering the dat polypeptide, to induce an antibody and/or T cell  
 CC immune response to protect against disease. This method is especially  
 CC useful for preventing bacterial infections (especially S. aureus) from  
 CC surgical implants e.g. pacemakers. The dat polypeptide may be used as a  
 CC wound treatment to prevent adhesion of bacteria to matrix proteins. Dat  
 CC antibodies are also useful for inducing an immune response to immunise  
 CC and prevent disease, and for isolating dat clones or purifying the  
 CC polypeptide by affinity chromatography. Diseases diagnosed, prevented or  
 CC treated include: infections of the upper and lower respiratory tract,  
 CC cardiac diseases, gastrointestinal disorders, central nervous system  
 CC disorders, eye diseases, kidney and urinary tract disorders, skin  
 CC disorders, and bone and joint disorders. Dat polypeptides are useful for  
 CC assaying genetic variation, and for mapping genes to chromosomes,  
 CC allowing gene inheritance to be studied through linkage analysis.  
 CC  
 XX Sequence 282 AA;

Query Match 16.4%; Score 290; DB 20; Length 282;  
 Best Local Similarity 26.7%; Pred. No. 1.6e-20;  
 Matches 77; Conservative 55; Mismatches 124; Indels 32; Gaps 6;  
 QY 49 WIEGYYLPAEAKISIFPTGSHDLTYVAHWGNIIFGLDRLDGLASAKLRIDAG 108  
 DB 5 FLNGEYVSPSEAKSYNDRGVYFGDGIETRYNGKILFTTEHYERLBSANIGIDLN 64  
 QY 109 YSKDELAETKCYMSQLESFVNLVTYRGYGRKGEKDLSKLTQVYIYAIPLYLWAP 168  
 DB 65 YSVEELIELSKRLVDMNIGETGAIYIQATRGVAERNHSFPPEVPAIVATKSYDPYD 124  
 QY 169 PAEYJFTGTAIVPRHVRAGRNIVDPTIKNYQW-----GDLTAASFEEKDKGAR 217

DB 125 HLEN-----GVNGV--TVEDIRMLRCIDIKSLNLGNLAKEYAVKYNAVE 167  
 QY 218 TAILLDSDNCVAREPGFNVCYKDKL-ASPSN-ALPGITRTYVELADQNGIEATLRD 275  
 DB 168 A--TQHRGETVTEGSSSSNAYAIKDVYTHPINNYLNTIRIVIKIAEDYNIPKEET 225  
 QY 276 VTSRELYADELMAVTTAGGVPIINSIDGAVNGECPGLYVAIRDRE 323  
 DB 226 FTVDPLKMADEYIVSSTSAEVPVIRKLDGEPIINDGKVGPIITRQLQSGF 273

RESULT 10  
 AAY13559  
 ID AAY13559 standard; Protein; 289 AA.

AC AAY13559;

DT 30-JUL-1999 (first entry)

DE L. monocytogenes D-amino acid aminotransferase (dat) gene product.

KW Listeria; vaccine; auxotrophic; attenuated; mutation; dat gene; dal gene;  
 KW T-cell response; HIV-1 antigen; HIV; aminotransferase.

OS Listeria monocytogenes.

PN W0925376-A1.

PD 27-MAY-1999.

PF 13-NOV-1998; 98MO-US24357.

PR 18-NOV-1997; 97US-0972902.

PA (UYPE-) UNIV PENNSYLVANIA.

PI Frankel FR, Portnoy DA;

DR WPI: 1999-337876/28.

DR N-PSDB: AAX55766.

XX Attenuated Listeria strain which is auxotrophic for D-alanine

PT synthesis

PS Disclosure; Fig 3A-C; 67pp; English.

CC The invention provides a vaccine comprising an auxotrophic attenuated  
 CC strain of Listeria which expresses an antigen, where the strain comprises  
 CC a mutation in at least one gene essential for growth of the Listeria  
 CC especially that the strain is auxotrophic for D-alanine. The gene is  
 CC selected from the Listeria dat and dal genes. The vaccine can be used in  
 CC mammals to stimulate T-cell responses to an antigen, especially an HIV-1  
 CC antigen. This is useful in creating effective vaccines against  
 CC HIV. The attenuated Listeria strain can be used as a bacterial vaccine  
 CC vector to raise an immunoresponse in e.g. immunocompromised patient or a  
 CC pregnant patient. The present sequence represents the L. monocytogenes  
 CC D-amino acid aminotransferase (dat) gene product.  
 CC  
 XX Sequence 289 AA;

Query Match 16.4%; Score 289; DB 20; Length 289;  
 Best Local Similarity 28.9%; Pred. No. 2.1e-20;  
 Matches 83; Conservative 51; Mismatches 123; Indels 30; Gaps 6;

QY 50 IBEYIPLAEBAKISIFDTGSHDLTYVAHWGNIIFGLDRLDGLASAKLRIDAG 109  
 DB 5 VNNHLVEREDATYDIEDRGYQEGDGYEVVRLYNGEFYFNHIDRLVSAKRIDLYIPY 64  
 QY 110 SKDELAETKCYMSQLESFVNLVTYRGYGRKGEKDLSKLTQVYIYAIPLYLWAP 169  
 DB 65 SKRELLELEKLVAAENNINNTGNVYLDYTRGVQNPKN-----HVIIPD--DFPL 109  
 QY 170 AEQIFGTGTAIVPRHVRAGRNIVDPTIKNYQW-----GDLTAASFEEKDKGAR 218

DB 110 EGVLTAAAREVPRNERQVTEGTAITEEDVRLKCDIKSLNLCNLIANK-KAHQNMAL 168  
 QY 219 AILLDSNCVAEGPGFNVCIVKDKLASPSRN--ALPGITRKTFEELADMGIEATLNDV 276  
 DB 169 AILHREGQ-VTECSASNVSIKDKGLMTWHAADNLILNGITROYIIVAKKNGIPVKEADF 227  
 QY 277 TSRELYADELMAVTTAGVTPINSIDGAVNGEPPGLVAIRDRF 323  
 DB 228 TLTLDRADEVFTISSTTETPTTHIDGVQADGKRGPTTAQLHOYF 274

## RESULT 11

ABBA9432 standard; Protein; 289 AA.

AC ABB49432;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes protein #2136.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 vitamin B12; bacterial infection; disease.

OS Listeria monocytogenes.

PN W0200177335-A2.

PD 18-OCT-2001.

PE 11-APR-2001; 2001MO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

PA (INSP) INST PASTEUR.

PI Buchrieser C, Frangoul L, Couve E, Rusnick C, Esli H, Dehoux P,  
 Dusarquet O, Chetouani F, Nedjar H, Glaser P, Kunst F, Cossart P,  
 Daniels J, Goebel W, Kretz J, Kuhn M, Ng E, Vazquez-Boland JA,  
 Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,  
 Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 Madueno E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J,  
 Rose M, Voss H;

DR WPI: 2002-010914/01.

PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 and prevention of Listeria and related bacterial infections, and  
 related polypeptides

PS Claim 6; SEQ ID NO 2137; 192pp; French.

CC The present invention relates to the genome sequence of Listeria  
 CC monocytogenes BGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in L.  
 CC monocytogenes and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies. Identification of L. monocytogenes and related organisms,  
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication,  
 CC and modulate L. monocytogenes-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by L.  
 CC monocytogenes and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIP0  
 CC at ftp.wip0.int/pub/published\_pct\_sequences.  
 CC

SO Sequence 289 AA:

Query Match 16.4%; Score 289; DB 23; Length 289;  
 Best Local Similarity 28.9%; Pred. No. 2, 1e-20;  
 Matches 83; Conservative 51; Mismatches 123; Indels 30; Gaps 6;

QY 50 IEGEYLPAREAKISTDTPFGSHDLTYVAHWGNTFRGLDHLDDGASKLRIDGY 109  
 DB 5 VNNHLERDADAVYDIEDRGYQFGDGYEVRRLYNGFFYYNEHIDRLYASAKIDLVPY 64  
 QY 110 SKDELAETTRKCCVSQMSLESFNLVTGKRGKRGKSLKTHQVYIATPIYMAEP 169  
 DB 65 SKDELLELEKLVAKENITGVYIATVTRGVONPRN-----HVIPID--DEPL 109  
 QY 170 ARIQFGTTAIVPRHVRAGRNTPPTIKYQW-----GDLTAAFEAKDRCART 218  
 DB 110 EGVLTAAAREVPRNERQVTEGTAITEEDVRLKCDIKSLNLCNLIANK-KAHQNMAL 168  
 QY 219 AILLDSNCVAEGPGFNVCIVKDKLASPSRN--ALPGITRKTFEELADMGIEATLNDV 276  
 DB 169 AILHREGQ-VTECSASNVSIKDKGLMTWHAADNLILNGITROYIIVAKKNGIPVKEADF 227  
 QY 277 TSRELYADELMAVTTAGVTPINSIDGAVNGEPPGLVAIRDRF 323  
 DB 228 TLTLDRADEVFTISSTTETPTTHIDGVQADGKRGPTTAQLHOYF 274

## RESULT 12

AAV13560 standard; Protein; 282 AA.

AC AAV13560;

DT 30-JUL-1999 (first entry)

DE S. haemolyticus D-amino acid aminotransferase.

KW Listeria; vaccine; auxotrophic; attenuated; mutation; dat gene; dal gene;  
 T-cell response; HIV-1 antigen; HIV; aminotransferase.

OS Staphylococcus haemolyticus.

PN W09925376-A1.

PD 27-MAY-1999.

PE 13-NOV-1998; 98MO-US24357.

PR 18-NOV-1997; 97US-0972902.

PA (UYPE-) UNITV PENNSYLVANIA.

PI Frankel FR, Portnoy DA;

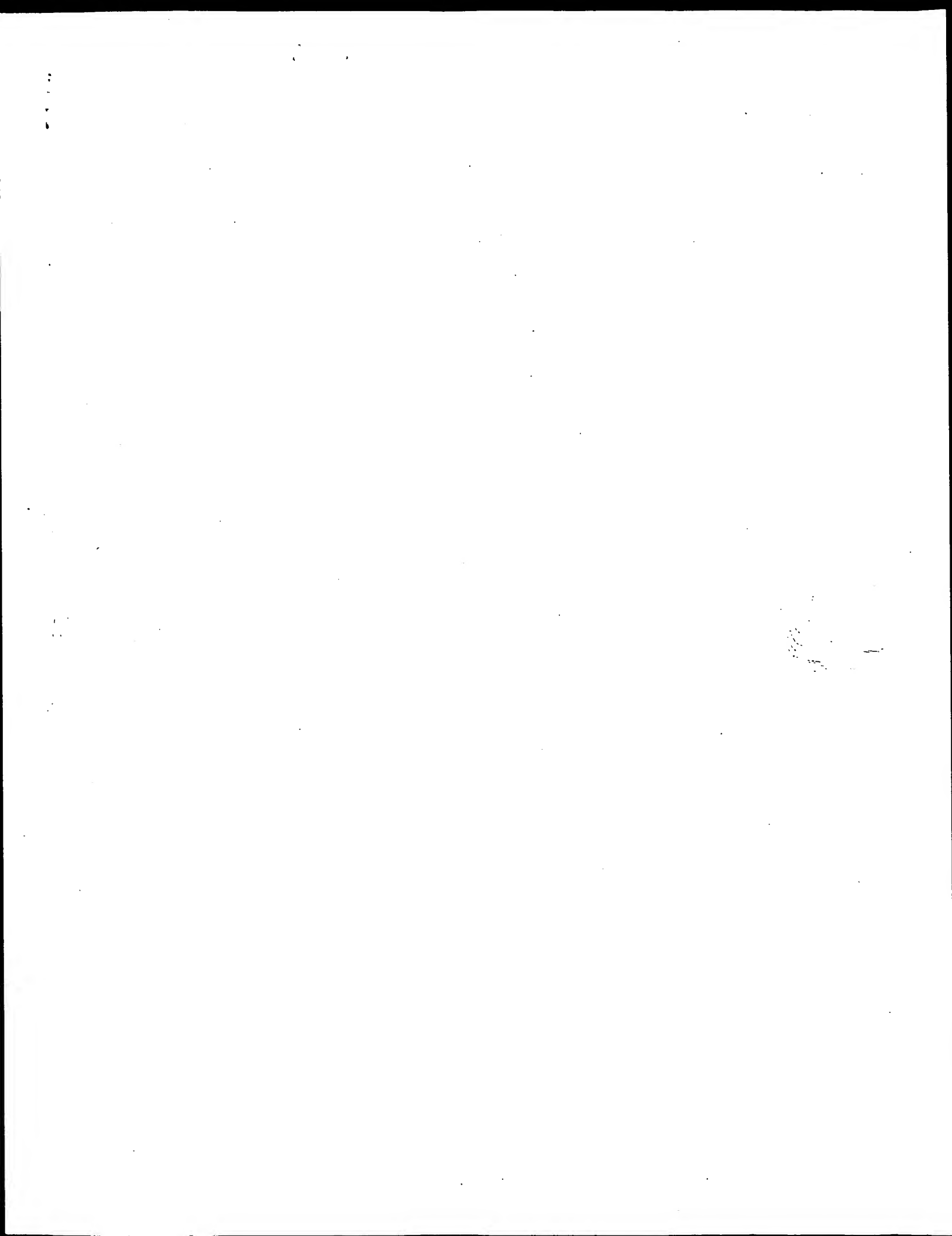
DR WPI: 1999-337876/28.

PT Attenuated Listeria strain which is auxotrophic for D-alanine  
 PT synthesis  
 PS Disclosure; Fig 4A-B; 67pp; English.

CC The invention provides a vaccine comprising an auxotrophic attenuated  
 CC strain of Listeria which expresses an antigen, where the strain comprises  
 CC a mutation in at least one gene essential for growth of the Listeria  
 CC especially that the strain is auxotrophic for D-alanine. The gene is  
 CC selected from the Listeria dat and dal genes. The vaccine can be used in  
 CC mammals to stimulate T-cell responses to an antigen, especially an HIV-1  
 CC antigen. This is useful in creating effective vaccines especially against  
 CC HIV. The attenuated Listeria strain can be used as a bacterial vaccine  
 CC vector to raise an immunoresponse in e.g. immunocompromised patient or a  
 CC pregnant patient. The present sequence represents a deduced D-amino acid  
 CC aminotransferase of S. haemolyticus.  
 CC









GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:35:56 ; Search time 19 Seconds

(without alignments)  
1715.240 Million cell updates/sec

Title: US-10-067-291-1

Perfect score: 1764  
Sequence: 1 MFAISDGLGTSNLVAVERGAI.....RDRFWALMDPEGLIETIEX 339

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR\_73:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	407.5	23.1	290	2	E69366
2	385	21.8	288	2	G64425
3	319	18.1	330	2	F69057
4	317.5	18.0	311	2	C70463
5	304.5	17.3	307	2	A83021
6	303.5	17.2	308	1	A34082
7	302.5	17.1	309	2	AD0924
8	301	17.1	308	2	A10474
9	298.5	16.9	309	2	H91216
10	298.5	16.9	309	2	A86063
11	298	16.9	308	2	B84362
12	296.5	16.8	309	1	XNDCV
13	290	16.4	289	2	B89960
14	289	16.4	282	2	AC1277
15	278.5	15.8	282	2	E69829
16	278.5	15.8	283	2	A31422
17	277.5	15.7	319	2	G82374
18	273.5	15.5	296	2	G87611
19	273	15.5	293	2	AB5555
20	264	15.0	289	2	AC1640
21	260	14.7	283	2	C84001
22	259	14.7	288	2	AD2885
23	259	14.7	319	2	A97661
24	253	14.3	304	2	D81445
25	244.5	13.9	285	2	D84197
26	244	13.8	288	2	A13217
27	240	13.6	287	2	D87465
28	238.5	13.5	283	2	G95267
29	212	12.0	290	2	B97774

## ALIGNMENTS

30	212	12.0	305	2	S75945	hypothetical prote
31	210.5	11.9	273	2	C72328	transaminase B hom
32	210	11.9	290	2	C71701	branched-chain ami
33	205.5	11.6	280	2	E96997	D-amino acid amino
34	204.5	11.6	271	2	G83275	4-amino-4-deoxycho
35	164.5	9.3	410	2	T46034	branched-chain-ami
36	160.5	9.1	295	2	D83661	4-amino-4-deoxycho
37	157	8.9	358	2	D89823	hypothetical prote
38	153	8.7	293	2	C37854	aminodeoxychorisna
39	144	8.2	384	2	T00625	branched-chain ami
40	141	8.0	1421	2	T05892	hypothetical prote
41	139.5	7.9	318	2	T00626	branched-chain ami
42	137	7.8	367	2	C96537	hypothetical prote
43	136	7.7	293	2	AG2883	branched-chain ami
44	136	7.7	342	2	E97465	branched-chain ami
45	132.5	7.5	280	2	D97174	enzyme of ILVE/PAB

## RESULT 1

branched-chain amino acid aminotransferase (ILVE) homolog - Archaeoglobus fulgidus

C/Species: Archaeoglobus fulgidus

C/Date: 05-Dec-1997 #sequence\_rev1510n 05-Dec-1997 #text\_change 18-Jun-1999

C/Accession: E69366

R/Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

Glocke, A.; Zhou, L.; Overbeek, R.; Cooney, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A/Authors: Usterbach, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kahne, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A/Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A/Reference number: A69250; MID:98049343; PMID:9389475

A/Accession: E69366

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-290 <KLE>

A/Cross-references: GB:AE001039; GB:AE000782; NID:92689362; PIDN:AAB90305.1; PID:9264

C/Superfamily: branched-chain-amino-acid aminotransferase

Query Match	Score	23.1%	DB 2:	Length	290;
Best Local Similarity	33.1%	Pred. No.	2.7e-25;		
Matches	100;	Conservative	60;	Mismatches	111;
				Indels	31;
				Gaps	7;
QY	49	WIEGEYLPAAERAKISIDTGFHSDLTFTYVAHWGNIIFRLGDHLRLDGLASKRLDAG	108		
DB	5	YMDGEFVDENAKVSIFDHGFLYGDVFEIGIRAVNGVFLKLEHIDRLYDSAKAIDLEIP	64		
QY	109	YSKDELAETIKCVSMQSLRESFVNLVTGTYG-----KRGK-----KDLSTHQ	155		
DB	65	ITKEEFMEIETLRKNNLRDARYRPTVTRIGLGLDPRCPNSIIVTKPWGKLYGD	124		
QY	156	VYIYAIPYMAFPFAEOIFGFTAIVRHVRAGNVTPTKKNYOMGDLTAASPEARDG	215		
DB	125	LY-----EK-GITAITVA-VRRNSDALPPIKSLNLTNLIKLENAK	168		
QY	216	ARTAILDSDNCAVAGPGFNVCIYKDGKIAS-PSRNALPGITRTKTVELADOMGIEATLR	274		
DB	169	GDEAIFIDRNGYVSGSDNIFVVKNGAITPPTINNLRGTRAVAIETINRLGIPPKET	228		
QY	275	DVTSRELYADDELAAVYTAGVYTPINSIDGAVNGEPPGLTVAIRDRFALMDPEG-PL	333		
DB	229	NIGYLDYADVEVVTGTAABIAYIVIDGRKIDGRKGETIRKLEMEFSKLTSEGVPI	288		
QY	334	IE 335			
DB	289	YE 290			

## RESULT 2

G64425

branched-chain amino acid aminotransferase - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C>Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 21-Jul-2000  
 C:Accession: G64425  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 Iron, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurlst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
 C.A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A:Reference number: A64300; PMID:96337999; PMID:8688087  
 A:Accession: G64425  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-288 <BDU>  
 A:Cross-references: GB:D67543; GB:L77117; NID:g1591663; PIDN:AAB99010.1; PID:g1591667; T  
 C:Genetics:  
 A:Map position: REV939094-938228  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 21.8%; Score 385; DB 2; Length 288;  
 Best Local Similarity 31.9%; Pred. No. 1.8e-23;  
 Matches 90; Conservative 58; Mismatches 116; Indels 18; Gaps 5;

OY 49 WIEGYVLAEEAKISIFDPTGCHSDLTVTVAHVHGNIFRGDHDRLDQASKRLDAG 108  
 Db 4 YLNGKVFDEKDKAVSFPHGLYGDVVEGIRADGVVFKLKEHIDRLYDASKSLCIDIP 63  
 OY 109 YSKDELAETTKCVMSQLRSEFVNLVTRGYG-----KRKGEKDLKLTQVYTAIPY 163  
 Db 64 LTKEMIDVLETLKVNLRDAVIRLVYTRGVGDLGIDPRKCGP-----TTCIAIP 116  
 OY 164 LMAFPAPQIGSTAIYVRHVRAGKRVTPITKNYQMGDITLAAFEAKDRGARTAILD 223  
 Db 117 --MPRLGEGDIRALTY-SVRLRPVDVLPNPAVSLNLTNSVLKIQANTYAGVDEAFILD 172  
 OY 224 SDNCVAFEPGFVNCIVKDGKLASPS--RNALPGITRKVPFLADOMGEATRLRYTSREL 281  
 Db 173 DKGFVEBTGDNIFVKNKGLKTPPYVSIKGTITRDVYIKLAEGLVEEPLTLHDL 232  
 OY 282 YDADELMAVTTAGVTPINSLDGEAVNGEPPLTVAIRDRF 323  
 Db 233 YTADELFTGTAAEIVPVEFIDGRVINNKQVGEITKKKEKF 274

## RESULT 3

#969057  
 branched-chain amino-acid aminotransferase - Methanobacterium thermoautotrophicum (strai

C:Species: Methanobacterium thermoautotrophicum  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 18-Jun-1999

C:Accession: F69057  
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
 ; Qiu, D.; Spadotora, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.  
 K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
 J. Bacteriol. 179, 7135-7155, 1997

A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct  
 A:Reference number: A69000; PMID:98037514; PMID:9371463  
 A:Accession: F69057  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-330 <MTH>  
 A:Cross-references: GB:AE000905; GB:AE000666; NID:g2622541; PIDN:AAB85907.1; PID:g262254  
 A:Experimental source: strain Delta H  
 C:Genetics:  
 A:Gene: MTH1430  
 A:Start codon: TTG  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 18.1%; Score 319; DB 2; Length 330;  
 Best Local Similarity 30.2%; Pred. No. 4.5e-18;  
 Matches 93; Conservative 50; Mismatches 121; Indels 44; Gaps 9;

OY 44 AGVAMIEGEYLPAAEKISIFDTGFGHSDLTVTVAHVH-----GNT 86

Db 29 ASGKIWLNGEWEWEATV-----HVLSHVHYGSSVFEGINCRNSGSAI 75  
 OY 87 FRLDGHDRLDQASKRLDAGSKDLAETTKCVMSQLRSEFVNLVTRGYGRKE 146  
 Db 76 FRLEHKKRLPDSAKIRMDIPTQEQICAIYETVENGLEBCYIRPVFRFY-----GE 131  
 OY 147 KDLSKLHVQVYIAPYLA--APPAAEQI-DGTTAIYVRHVRAGKRVTPDPTIK--NYQM 201  
 Db 132 MGIVPVCVPDVAVAAEMGAYIGAELEGVYDAGYSTW--RRAPRTMPTMAAGNYLM 190  
 OY 202 GDLTAASFEAKDRGARTAILLSDNCVAEGPGFNVCIVKGLASP--SRNALPGITRTY 259  
 Db 191 SOL--AKMEAVRHGYDEAIMLDYHGYISGSGENIFLVSEGIYTPPVSSLLRGITRDS 248  
 OY 260 VFEELADMGTEATRLROYTSRELVDADDELMATVTAAGVTPINSLDGEAVNGEPPLTVAI 319  
 Db 249 VIKIARTGVTVEEPTREMLYIADAFPTGTAAETTPIRSYDGIETGAGRGPVTKLL 308  
 OY 320 RDRFALM 327  
 Db 309 QDEFFRII 316

## RESULT 4

#70463  
 branched-chain amino acid aminotransferase - Aquifex aeolicus

C:Species: Aquifex aeolicus  
 C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Jun-1999

C:Accession: C70463  
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
 V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; PMID:98196666; PMID:9537320  
 A:Accession: C70463  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-311 <AOQ>  
 A:Cross-references: GB:AE000762; NID:g2984163; PIDN:AAO7697.1; PID:g2984164; GB:AE00  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: live  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 18.0%; Score 317.5; DB 2; Length 311;  
 Best Local Similarity 29.5%; Pred. No. 5.4e-18;  
 Matches 92; Conservative 46; Mismatches 119; Indels 55; Gaps 7;

OY 43 FAGVAMIEGEYLPAAEKISIFDTGFGHSDLTVT--VAHVHGN-----IFRLGDHDLRL 96  
 Db 5 FSDMFAFFEGKIVPVEAKISINTNSFHYGTAFEGIRAVNVEEEOQLYLFKHYERL 64  
 OY 97 LQASLRLDAGSKDELAEITKCVMSQLRSEFVNLVTRGYKRRGEKDLKLTQV 156  
 Db 65 LTNARCLFELNLSAELEITKEILRSEIRE-----DV 99  
 OY 157 YIYAIPLYLWAPPAAEQIFGTATVPRHVRAGR-----NTVPTIKNTOMGD----- 203  
 Db 100 YIRPAIVFDMLKLTPLIDYTPEDAIYLRFRYLDTSIGIRAKKSSRRNDNSIPSRW 159  
 OY 204 -----LTAASFEAKDRGARTAILLSDNCVAGPGFNVCIVKGLASPSR--ALP 253  
 Db 160 KVAGAVNSALKKTEALMSGYDEAILNSOGYVAGSGENIFIIKNGKATPPSPREHILE 219  
 OY 254 GTRKTVFE-LADMGIEATRLDYTSRELVDADDELMATVTAAGVTPINSLDGEAVNGEP 312  
 Db 220 GITRNAVITLKKELVEVEERPIARSELTYADEVFLGTAAEVPPVVEIDNRKIGNEI 279  
 OY 313 GPLTVAIRDRFW 324  
 Db 280 GITTKQLQGEFYF 291

## RESULT 5

A83021  
branched-chain amino acid transferase PA5013 [imported] - *Pseudomonas aeruginosa* (strain  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83021  
R:Stover, C.K.; Pham, X.O.; Ewlin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lam,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83021  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <STO>  
A:Cross-references: GB:AE004913; GB:AE004091; NID:g9951286; PIDN:AA08398.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: 116; PA5013  
C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 17.3%; Score 304.5; DB 2; Length 307;  
Best Local Similarity 30.0%; Pred. No. 6e-17;  
Matches 93; Conservative 47; Mismatches 117; Indels 53; Gaps 10;

QY 46 GVAMTEGEYLPAAERAKISIFDTGFGHSDLTYYVAHGN-----IFRLGDH 92  
DB 8 GVMTDGLVGMWDRATTHV-----LHITLHYGKGVGEGVAVAYTPPGTALFRQA 58  
QY 93 LDRLLDGSASRLDAGYSKDELAETKRCVMSQLESFVNLV--TRGYGR-KGEKD 148  
DB 59 TDRLEFDSAHIMNMQIPVSRDEINATRAVRENMLESAIYIPVYFSEGGMLRASG 117  
QY 149 LSKLTHOYITAIPIYLAFFPAEQIF-----GTAIVPRHVR---RAGNTYDPIKN 198  
DB 118 -----VHTIAMSVCWGYGEEALDQGIKVRITSSFTRHVNISMTRASNGA----- 164  
QY 199 YQMGDLTFASEAKDRCARATAILDSDNCVAFEGFVNCYIKDKLASPSRNA-LPGITR 257  
DB 165 --YNSMALDGAISGDEAMMDPEGVAVGSGENFIITKDVITTPETVTAQLNIGTR 222  
QY 258 KTFEVLADQMGIEATLRDVTRELYYADELMAVTTAGGVPINSLDGEAVNGEPGLTV 317  
DB 223 NITLTLAEHGFKEKRTIDREYIADAEAFETGTALEVPIRVEDGKIGACGRGPVTE 282  
QY 318 AIRDRFALM 327  
DB 283 KIQKAYFDLV 292

## RESULT 6

A34082  
branched-chain-amino-acid transaminase (EC 2.6.1.44) - *Salmonella typhimurium*  
N:Alternate names: branched-chain-amino-acid aminotransferase; transaminase B  
C:Species: *Salmonella typhimurium*  
C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 30-Apr-1999  
C:Accession: A34082; A05077  
R:Feild, M.J.; Nguyen, D.C.; Armstrong, F.B.  
Biochemistry 28, 5306-5310, 1989  
A:Title: Amino acid sequence of *Salmonella typhimurium* branched-chain amino acid aminot  
A:Reference number: A34082; MUID:89352621; PMID:2669973  
A:Accession: A34082  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-308 <FEI>  
R:Randall, R.R.; Wallis, M.H.; Young, G.J.; Armstrong, F.B.  
Fed. Proc. 38, 3234, 1979  
A:Title: N-terminal sequence of branched-chain amino acid aminotransferase.  
A:Reference number: A05077  
A:Accession: A05077  
A:Molecule type: protein

## A:Residues: 1-5 &lt;RAN&gt;

C:Function:  
A:Description: catalyzes the reversible transamination of alpha-amino groups from leu  
C:Pathway: branched-chain amino acid biosynthesis  
C:Superfamily: branched-chain-amino-acid aminotransferase  
C:Keywords: aminotransferase; branched-chain amino acid biosynthesis; phosphoprotein;  
F:159/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 17.2%; Score 303.5; DB 1; Length 308;  
Best Local Similarity 31.6%; Pred. No. 7.2e-17;  
Matches 91; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

QY 49 WIEGEYLPAAERAKISIFDTGFGHSDLTYYVAHGN-----IFRLGDHRLDGSASRL 104  
DB 9 WFGKGVNEDAKVHVMASHALHYGTSVEFGICRYDSHKPQVFRHREHQRDLSAKIYR 68  
QY 105 LDAGYSKDELAETKRCVMSQLESFVNLVTRGKRGKGEKDLSTLHOYIATPIYL 164  
DB 69 FVYSQSIDELMEACRDVIRKNNLTSAIYRPLVFG--DVGMGVNPPPGYTTDYITAAFPW- 126  
QY 165 WAFPPAEQI-FGTTATVPRHVRAGNTVDPYIK--NYQMGDLTFASEAKDRCARAT 220  
DB 127 GAVLGAELDQGIDAVVSSM-NRAAPNTIPTAAKAGNYLSILVGS--EARRHGYQEG 183  
QY 221 LDDSDNCVAFEGFVNCYIKDKLASP--SRNALPGITRKTFELADQMGIEATLRDVT 278  
DB 184 ALDVNGYISGAGENLFEKRGVLPPTSSALPEITDAILIKAKELGIEVRDQVLSR 243  
QY 279 RELYDDELMAVTTAGGVPINSLDGEAVNGEPGLTVAIRDRFAL 326  
DB 244 ESILYLADEVFMSTALAEITPVRSVDIQYGEKGCVPYTRIQQAIFGL 291

## RESULT 7

branched-chain amino-acid aminotransferase [imported] - *Salmonella enterica* subsp. en  
AD0924  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AD0924  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AD0924  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-309 <PAR>  
A:Cross-references: GB:AL513862; PIDN:CAD09414.1; PID:g16504532; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY3654  
C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 17.1%; Score 302.5; DB 2; Length 309;  
Best Local Similarity 31.6%; Pred. No. 8.7e-17;  
Matches 91; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

QY 49 WIEGEYLPAAERAKISIFDTGFGHSDLTYYVAHGN-----IFRLGDHRLDGSASRL 104  
DB 10 WFGKGVNEDAKVHVMASHALHYGTSVEFGICRYDSHKPQVFRHREHQRDLSAKIYR 69  
QY 105 LDAGYSKDELAETKRCVMSQLESFVNLVTRGKRGKGEKDLSTLHOYIATPIYL 164  
DB 70 FVYSQSIDELMEACRDVIRKNNLTSAIYRPLVFG--DVGMGVNPPPGYTTDYITAAFPW- 127  
QY 165 WAFPPAEQI-FGTTATVPRHVRAGNTVDPYIK--NYQMGDLTFASEAKDRCARAT 220  
DB 128 GAVLGAELDQGIDAVVSSM-NRAAPNTIPTAAKAGNYLSILVGS--EARRHGYQEG 184  
QY 221 LDDSDNCVAFEGFVNCYIKDKLASP--SRNALPGITRKTFELADQMGIEATLRDVT 278

Db 185 ALDVNGYISBAGENLEFEVKGGLVFTPTSSALPGITRDALIKLAKELGIEVEQVLSR 244  
 QY 279 RELYDADELMAYTTAGVTPINSLDGEAVNGEPLTVAIRDFMAL 326  
 Db 245 ESYLADEVMGSGTAAETTPVRSVDGIQVGEGRGCPVTKRIQAAFFGL 292

## RESULT 8

A10474

branched-chain amino-acid transaminase (EC 2.6.1.42) [imported] - Yersinia pestis (strain

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001

C:Accession: A10474

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tildall, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarrara, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: A10474

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-308 &lt;KUR&gt;

A:Cross-references: GB:AL590842; PIDN:CAC93365.1; PID:g15981811; GSPDB:GN00175

C:Genetics:

A:Gene: 11VE

C:Superfamily: branched-chain-amino-acid aminotransferase

C:Keyfamily: aminotransferase

Query Match

Best Local Similarity 17.1%; Score 301; DB 2; Length 308;  
 Best Local Similarity 30.2%; Pred. No. 1.1e-16;  
 Matches 92; Conservative 46; Mismatches 141; Indels 26; Gaps 9;

QY 49 WIEGEYLPAAEAKISIFDTGFGHSDLYTVAVHNGN----IFRLGDHLDRLLDGASKLR 104  
 Db 9 WFNEMVPMAEKAVHVMHSHLHGTSTVEGRTYSHKGVYRHRHMRRLHDSAKIYR 68  
 QY 105 LDAGYSKDELAETTKCVSMQSLRESFVNLVTRGKRGKEDLSKLTQVYIAYPYL 164  
 Db 69 MPVQSIDELMEACRDVIRKNNLTSVIRPLIFVG-DVGMGVNPPAGYSTVYIAAFPM- 126  
 QY 165 WAFPPAEQI-FGTTAIVPRHVRAGRNVDPTIK---NYQWGDLTAAEFKDRGARTAI 220  
 Db 127 GATLGAELAEQIDAVSSW-NRAAPNTIPTAAKAGNTLSLVGS--EARRHGYOEGI 183  
 QY 221 LLDSDNCVAEGPGFNVCIYKDGKIASP--SRNALPGITRKTVEFLADQMGIEATLDVYS 278  
 Db 184 ALDVNGYISBAGENLEFEVKGGLVFTPTSSALPGITRDALIKLAKMGLEVEQVLSR 243  
 QY 279 RELYDADELMAYTTAGVTPINSLDGEAVNGEPLTVAIRDRF-----WALM 327  
 Db 244 ESYLADEVMGSGTAAETTPVRSVDGIQVGEGRGCPVTKRIQAAFFGLTGTEDKMWGL 303  
 QY 328 DEGP 332  
 Db 304 DQVNP 308

## RESULT 9

H91216

branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157:H7

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C:Accession: H91216

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasaavara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H91216

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-309 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA938127.1; PID:g13364119; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECS4704  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 16.9%; Score 298.5; DB 2; Length 309;  
 Best Local Similarity 30.9%; Pred. No. 1.8e-16;  
 Matches 89; Conservative 49; Mismatches 135; Indels 15; Gaps 8;

QY 49 WIEGEYLPAAEAKISIFDTGFGHSDLYTVAVHNGN----IFRLGDHLDRLLDGASKLR 104  
 Db 10 WFNEMVPMAEKAVHVMHSHLHGTSTVEGRTYSHKGVYRHRHMRRLHDSAKIYR 69  
 QY 105 LDAGYSKDELAETTKCVSMQSLRESFVNLVTRGKRGKEDLSKLTQVYIAYPYL 164  
 Db 70 MPVQSIDELMEACRDVIRKNNLTSVIRPLIFVG-DVGMGVNPPAGYSTVYIAAFPM- 127  
 QY 165 WAFPPAEQI-FGTTAIVPRHVRAGRNVDPTIK---NYQWGDLTAAEFKDRGARTAI 220  
 Db 128 GATLGAELAEQIDAVSSW-NRAAPNTIPTAAKAGNTLSLVGS--EARRHGYOEGI 184  
 QY 221 LLDSDNCVAEGPGFNVCIYKDGKIASP--SRNALPGITRKTVEFLADQMGIEATLDVYS 278  
 Db 185 ALDVNGYISBAGENLEFEVKGGLVFTPTSSALPGITRDALIKLAKELGIEVEQVLSR 244  
 QY 279 RELYDADELMAYTTAGVTPINSLDGEAVNGEPLTVAIRDFMAL 326  
 Db 245 ESYLADEVMGSGTAAETTPVRSVDGIQVGEGRGCPVTKRIQAAFFGL 292

RESULT 10  
 A86063  
 branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: A86063  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamowski, K.; Apoda  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Accession: A86063  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-309 <STO>  
 A:Cross-references: GB:AE005174; NID:g12518628; PIDN:AAG58965.1; GSPDB:GN00145; UMG:  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: 11VE  
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 16.9%; Score 298.5; DB 2; Length 309;  
 Best Local Similarity 30.9%; Pred. No. 1.8e-16;  
 Matches 89; Conservative 49; Mismatches 135; Indels 15; Gaps 8;

QY 49 WIEGEYLPAAEAKISIFDTGFGHSDLYTVAVHNGN----IFRLGDHLDRLLDGASKLR 104  
 Db 10 WFNEMVPMAEKAVHVMHSHLHGTSTVEGRTYSHKGVYRHRHMRRLHDSAKIYR 69  
 QY 105 LDAGYSKDELAETTKCVSMQSLRESFVNLVTRGKRGKEDLSKLTQVYIAYPYL 164  
 Db 70 MPVQSIDELMEACRDVIRKNNLTSVIRPLIFVG-DVGMGVNPPAGYSTVYIAAFPM- 127  
 QY 165 WAFPPAEQI-FGTTAIVPRHVRAGRNVDPTIK---NYQWGDLTAAEFKDRGARTAI 220  
 Db 128 GATLGAELAEQIDAVSSW-NRAAPNTIPTAAKAGNTLSLVGS--EARRHGYOEGI 184  
 QY 221 LLDSDNCVAEGPGFNVCIYKDGKIASP--SRNALPGITRKTVEFLADQMGIEATLDVYS 278  
 Db 185 ALDVNGYISBAGENLEFEVKGGLVFTPTSSALPGITRDALIKLAKELGIEVEQVLSR 244





```

Db 116 TVKKPBOEQAYGAATIDEDIRML--KCDIKSINILYNWTKO--RAYAGAFBAILL- 169
QY 224 SDNCVAEGPGFNVCIVDGLAS--PSRNALPGITRTKVFELADOMGIEATLRDVTISREL 281
Db 170 RDGVVTEGTSNMYAVANGTVRTHPANRLJLNGITRMIIGLIEKNGIKLDETPVSEEL 229
QY 282 YDADELMAYTTAGVTPINSIDGAVNGEGEPULTVAIRDRF 323
Db 230 KQAEIFISSTAEIIPVTLIDGOSISGKPGPYTKOLQAAF 271

```

Search completed: April 17, 2003, 17:39:55  
 Job time : 21 secs

200-1



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:29:31 ; Search time 25 Seconds

(without alignments)  
562.419 Million cell updates/sec

Title: US-10-067-291-1

Perfect score: 1764  
Sequence: 1 MIALSDLGTSNLVAVPGAI.....RDRFWALMDEPGPLETIEY 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	407.5	23.1	290 1	ILVE_ARCFU 029329 archaeoglob
2	385	21.8	288 1	ILVE_METUA 029814 methanococ
3	319	18.1	306 1	ILVE_METTH 027481 methanobact
4	317.5	18.0	311 1	ILVE_AQUAE 067733 aquifex aeo
5	309.5	17.5	283 1	P54692 bacillus 11
6	304.5	17.3	307 1	ILVE_PSEAE 086428 pseudomonas
7	302.5	17.1	308 1	ILVE_SALTY 151568 salmonella
8	298.5	16.9	308 1	ILVE_ECOLI 00510 escherichia
9	289	16.4	289 1	DAAA_LISMO 085046 listeria mo
10	283	16.0	282 1	DAAA_STARA 054694 staphylococ
11	278.5	15.8	282 1	DAAA_BACSP 019938 bacillus sp
12	278.5	15.8	282 1	DAAA_BACSU 007597 bacillus su
13	264	15.0	289 1	DAAA_LISIN 092590 listeria in
14	250	14.2	283 1	DAAA_BACSH 054693 bacillus sp
15	212	12.0	290 1	ILVE_RICCN 092126 rickettsia
16	212	12.0	305 1	ILVE_SYNY3 P54691 synechocyst
17	210.5	11.9	273 1	ILVE_THEMA P74921 thermotoga
18	210	11.9	290 1	ILVE_RICPR 005970 rickettsia
19	153	8.7	293 1	PABC_BACSU 028821 bacillus su
20	120.5	6.8	368 1	ILVE_MYCTU 010399 mycobacteri
21	119.5	6.8	356 1	YBGE_BACSU 031461 bacillus su
22	116	6.6	384 1	PABC_HUMAN P54687 homo sapien
23	114.5	6.5	267 1	PABC_VIBCH 094910 vibrio chol
24	113.5	6.4	368 1	ILVE_MYCLE 032954 mycobacteri
25	110.5	6.3	386 1	BCAT_MOUSE P24288 mus musculi
26	109.5	6.2	271 1	PABC_ECOLI 028305 escherichia
27	109.5	6.2	271 1	PABC_VIBHA 056693 vibrio harv
28	102.5	5.8	362 1	ILVE_BACSU P39576 bacillus su
29	100.5	5.7	362 1	ILVE_STRCO 086505 streptomyc
30	100.5	5.7	411 1	BCAT_RAT P54690 rattus norv
31	100	5.7	874 1	SYA_PASMU P57933 pasteurella
32	99.5	5.6	607 1	DNAK_STRPN P95829 streptococ
33	99.5	5.6	1473 1	TOP2_ARATH P30182 arabidopsis

34	99	5.6	638 1	DNAK_PRETO 0939f1 prevotella
35	98.5	5.6	807 1	AFSK_STRKR P44742 streptomyc
36	98	5.6	385 1	ACKA_MYCTU P66235 mycobacteri
37	96.5	5.5	610 1	DNAK_BACSH 069268 bacillus sp
38	95.5	5.4	393 1	BCAM_RAT 035854 rattus norv
39	95	5.4	379 1	YB29_BRUME 08ybn1 bruceella me
40	95	5.4	1076 1	CARB_HALER 08rs3 halomonas e
41	92	5.2	1081 1	CARB_RALSO 08x283 ralsstonia s
42	91.5	5.2	343 1	ILVE_HAEIN P54689 haemophilus
43	91.5	5.2	1014 1	BGAL_BACHD 09K66 bacillus ha
44	91	5.2	721 1	BBS2_HUMAN 09bxc9 homo sapien
45	91	5.2	3038 1	TRIO_HUMAN 075962 homo sapien

## ALIGNMENTS

RESULT 1  
ID ILVE\_ARCFU STANDARD: PRT: 290 AA.  
AC 029329;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)  
GN ILVE OR AF0933.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klek H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyprides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kitzness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,  
RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC - FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
CC - CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate -> 4-methyl-2-  
oxopentanoate + L-glutamate.  
CC - COFACTOR: PYRIDOXAL PHOSPHATE (By similarity).  
CC - PATHWAY: Valine and isoleucine biosynthesis.  
CC - SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
AMINOTRANSFERASES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL: AE001039; AAB90305.1; -  
CC HSSP: P19938; IAOG.  
DR TIGR: AF0933;  
DR InterPro: IPR001544; AminoTran\_4.  
DR Pfam: PF01063; AminoTran\_4; 1.  
DR PRODOM: PD001961; AminoTran\_4; 1.  
DR TIGRFAMs: TIGR01122; LIVE\_I; 1.  
DR PROSITE: PS00770; AA\_TRANSFER\_CLASS\_4; 1.

KW Transferrase; Aminotransferase; Branched-chain amino acid biosynthesis;  
 KM Pyridoxal phosphate; Complete proteome.  
 FT BINDING 150 PYRIDOXAL PHOSPHATE (POTENTIAL).  
 SQ SEQUENCE 280 AA; 32366 MW; F096FF8C3BA273 CRC64;

Query Match 23.1%; Score 407.5; DB 1; Length 280;  
 Best Local Similarity 33.1%; Pred. No. 2.2e-25;  
 Matches 100; Conservative 60; Mismatches 111; Indels 31; Gaps 7;

QY 49 WIEGELPAEAKISIFPDGFGSDLTYYVAHWGNIIFRLGDHDLRLDGASKLRLDAG 108  
 DB 5 YMDGEFVPEBEAKSIFPHGLGIDGVEEGRANRNGVFLKEHIDRLYDSAKAIDLEIFP 64  
 QY 109 YSKDELAIEIKKCVMSQLSRESFVNLTVRGVS-----KRKGE-----KDLSKLTHQ 155  
 DB 65 ITKEEFMEIETLRKNNLRDAVIRPIYTRGIDGLGIDPRCQNPISIVTRPKMKLYGD 124  
 QY 156 YVTAIPILMAFPPEAQJFTTAIVPRHVRAGRTVPTIKNQMGDLTAASEADRG 215  
 DB 125 LV-----EK--GLTAIVVA--VRRNSFDALPBNKISYLNMLAKTEANAKG 168  
 QY 216 ARTATLSDNCVAGPEPNCIYKDGKLAS-PSRNALPGITRKTYPFLADOMGTEATLR 274  
 DB 169 GDEALFLDRNGVSESGSDNIFVANKGAIPTPTNNLRGITREAVLEIINRLGIPERET 228  
 QY 275 DVTSEHLYDADELMAYTAGVTPINSIDGAVNGEPPLTVAIIRDFTALMDPEP-PL 333  
 DB 229 NIGLDVLTADAEVFTGTAELAPIVVDGRKIGDKGKELTRKMEFSLTSEGVPI 288  
 QY 334 IE 335  
 DB 289 YE 290

## RESULT 2

ILVE\_METHA STANDARD; PRT; 288 AA.  
 AC 058414;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)  
 DE (Transaminase B) (BCAT).  
 GN ILVE OR M01008.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 CX NCBI\_TaxID=2190;

RA SUTON G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Cocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Scott J.L., Geoghegan N.S.M., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,  
 RA Klenk H.P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.,  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RA jannaschii.";  
 RA Science 273:1058-1073(1996).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
 CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-  
 CC oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U67543; AAB99010.1; -  
 DR HSSP; P00510; IAG3.  
 DR TIGR; M01008; -  
 DR InterPro: IPR001544; AminoTran\_4.  
 DR Pfam: PF01063; aminoTran\_4; 1.  
 DR ProDom; PD001961; AminoTran\_4; 1.  
 DR TIGRfam; TIGR01122; ILVE\_1; 1.  
 DR PROSITE; PS00770; AA\_TRANSFR\_CLASS\_4; 1.  
 KM Transferrase; Aminotransferase; Branched-chain amino acid biosynthesis;  
 KM Pyridoxal phosphate; Complete proteome.  
 FT BINDING 146  
 SQ SEQUENCE 288 AA; 32096 MW; 033DBA17DFD89C8C CRC64;

Query Match 21.8%; Score 385; DB 1; Length 288;  
 Best Local Similarity 31.9%; Pred. No. 1.4e-23;  
 Matches 90; Conservative 58; Mismatches 116; Indels 18; Gaps 5;

QY 49 WIEGELPAEAKISIFPDGFGSDLTYYVAHWGNIIFRLGDHDLRLDGASKLRLDAG 108  
 DB 4 YLNGKFPDEKDAKVSFEDHGLTGDFEGRAYDGVFMLEKHIDRLYDSAKSLCIDIP 63  
 QY 109 YSKDELAIEIKKCVMSQLSRESFVNLTVRGVS-----KRKGEKDLKTHQVYIAYIPY 163  
 DB 64 LTRKEMIDVLETRVANNLRDAITIRLVTRGVGLDLPKRGKGP-----TIFCLATP- 116  
 QY 164 LMAFPPEAQJFTTAIVPRHVRAGRTVPTIKNQMGDLTAASEADRGARATLID 223  
 DB 117 ---MPRLGEGDGRATYV--SVRLPVDVLLNPAVKSINLYSLAKIOANYAGVDEAFLLD 172  
 QY 224 SDNCVAGPEPNCIYKDGKLASPS--RNALPGITRKTYPFLADOMGTEATLRDVTSRRL 281  
 DB 173 DKGFVEEGSDNIFVANKGAIPTPTNNLRGITREAVLEIINRLGIPERET 228  
 QY 282 YDADELMAYTAGVTPINSIDGAVNGEPPLTVAIIRDFTALMDPEP-PL 323  
 DB 233 YTADELFITTAELIYVFEIDRIVNNQVSGITKKLEKF 274

## RESULT 3

ILVE\_METHH STANDARD; PRT; 306 AA.  
 AC 027481;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Putative branched-chain amino acid aminotransferase (EC 2.6.1.42)  
 DE (Transaminase B) (BCAT).  
 GN ILVE OR MTH1430.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea: Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanothermobacter.  
 CX NCBI\_TaxID=187420;

RA SUTON G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Cocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Scott J.L., Geoghegan N.S.M., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kalne B.P., Borodovsky M.,  
 RA Klenk H.P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.,  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 RA jannaschii.";  
 RA Science 273:1058-1073(1996).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
 CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-  
 CC oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 CC -1- PATHWAY: valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AE000905; AAB85907.1; ALT\_INT.  
 DR HSP: P00510; IAB3.  
 DR InterPro: IPR001544; AminoTran\_4.  
 DR Pfam: PF01063; AminoTran\_4; 1.  
 DR ProDom: PD001961; AminoTran\_4; 1.  
 DR TIGRFAMs: TIGR01122; IIVE.1; 1.  
 DR PROSITE: PS00770; AA\_TRANSFERRER\_CLASS\_4; 1.  
 DR Transferrase; AminoTransferase; Branched-chain amino acid biosynthesis;  
 KW Pyridoxal phosphate; Complete proteome.  
 FT BINDING 159 PYRIDOXAL PHOSPHATE (POTENTIAL).  
 SQ SEQUENCE 306 AA; 34205 MW; 725611CAC4796F98 CRC64;  
 Query Match 18.1%; Score 319; DB 1; Length 306;  
 Best Local Similarity 30.2%; Pred. No. 2.7e-18; Indels 44; Gaps 9;  
 Matches 93; Conservative 50; Mismatches 121;  
 QY 44 AGCAWIEGEYLPAAEKISIFDTGFGHSDLYTVAAVHM-----GNI 86  
 DB 5 ASGKINLNGEVEVEENATV-----HLSHVHNGSSVEFGICRYRNSKGSAL 51  
 QY 87 FRLDGHLRLDLDGASKRLDAGYCKDELAETTKCVSNQIRSEFVNLTYRGCKRKE 146  
 DB 52 FRLSEHVRLPDSAKIYMDIPYQEQICDAIVETVRENGIECEYIRPVFRGY---GE 107  
 QY 147 KDLKRLHGVYIYAIPLYM-AEPPEOI-FGTATVPRHVRAGRNVDPTIK---NYQW 201  
 DB 108 MGVHPVNCVAVAAWMAKAYLDALEALGVVDAGVSTW-RRMAPNTPMAKAGGNTLN 166  
 QY 202 GDLTAASFEAKDRGARTAILLSDNCVAEGPGFVNCIVKDKLASP--SRNALPGITRTK 259  
 DB 167 SQL-AKKEAVRHHGDEALMDYHGYISEGSENIPLVSEGEIYTPPVSSLLRGITRDS 224  
 QY 260 VPELADONGIEATLADVTSRRLYDADELMAVTTAGGVTPINSLDGEAVNGNEPPLVVAI 319  
 DB 225 VIKIARTGVVHEPPTIREMLYLADEAFPTGTAAETPIRSVDGIEIGAGRGPPVKLL 284  
 QY 320 RDRFWALM 327  
 DB 285 QDEPRIT 292  
 RESULT 4  
 ID LIVE\_AQUAE STANDARD: PRT; 311 AA.  
 AC 067733;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42) (BCAT).  
 GN IIVE OR AQ\_1893.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;  
 OC Aquifex.  
 OX NCBI\_TaxID=63363;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;

RX MEDLINE-98196666; PubMed-9537320;  
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,  
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,  
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RL Nature 392:353-358 (1998).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
 CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AE000762; AAC07697.1; -  
 DR HSP: P00510; IAB3.  
 DR InterPro: IPR001544; AminoTran\_4.  
 DR Pfam: PF01063; AminoTran\_4; 1.  
 DR ProDom: PD001961; AminoTran\_4; 1.  
 DR TIGRFAMs: TIGR01122; IIVE.1; 1.  
 DR PROSITE: PS00770; AA\_TRANSFERRER\_CLASS\_4; 1.  
 DR Transferrase; AminoTransferase; Branched-chain amino acid biosynthesis;  
 KW Pyridoxal phosphate; Complete proteome.  
 FT BINDING 160 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 311 AA; 35763 MW; 0C02F2520947D647 CRC64;  
 Query Match 18.0%; Score 317.5; DB 1; Length 311;  
 Best Local Similarity 29.5%; Pred. No. 3.6e-18; Indels 55; Gaps 7;  
 Matches 92; Conservative 46; Mismatches 119;  
 QY 43 FAGCAWIEGEYLPAAEKISIFDTGFGHSDLYTVAAVHMGN-----IFRLGHLRL 96  
 DB 5 FSNDFPFEKIKYVPEAEKISIMTNSFHHGTAIFEDIRAYWNEEQRLILFAKHEYRL 64  
 QY 97 LDGASKRLDAGYCKDELAETTKCVSMQIRSEFVNLTYRGCKRKEKDLKLTROY 156  
 DB 65 LTNARCLFMELNYSAELEIETKELRKSEIR-----DV 99  
 QY 157 YIYAIPLYMAFPPEAOIFGTATVPRHVRAGR-----NTVDTIKNYQWD----- 203  
 DB 100 YIRPIAVFKDLKLPKLDITPPEIATYLRFGYLDTSKGIKAKVSSMRANDNSIPRW 159  
 QY 204 -----LTAASFEAKDRGARTAILLSDNCVAEGPGFVNCIVKDKLASPSR--ALP 253  
 DB 160 KVAQAVVNSALAKTEALMSGYDEAILNSQGYVAEGSENIPLIKKKAITSPNEHILE 219  
 QY 254 GITRKTYFE-LADONGIEATLADVTSRRLYDADELMAVTTAGGVTPINSLDGAVNGNEP 312  
 DB 220 GITRNAVITLLKRELVVERPIRSELYTADVELGTAAEVTVEVDINKKIGNGEI 279  
 QY 313 GPLVAVTRDRFW 324  
 DB 280 GPLTKQLQEFYF 291  
 RESULT 5  
 ID DAAA\_BACLI STANDARD: PRT; 283 AA.  
 AC P54682;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate  
 DE aminotransferase) (D-amino acid aminotransferase) (D-amino acid

DE transaminase) (DAAAT).  
GN DAT.  
OS Bacillus licheniformis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxId=1402;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 10716;  
RA MEDLINE=97157074; PubMed=9003455;  
RT Taylor P.P., Fotheringham I.G.;  
RT "Nucleotide sequence of the Bacillus licheniformis ATCC 10716 dat  
RT gene and comparison of the predicted amino acid sequence with those  
RT of other bacterial species."  
RL Biochim. Biophys. Acta 1350:38-40(1997).  
CC -1- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,  
CC GLUTAMATE, AMINOBUVATE, NORVALINE AND ASPARAGINE (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate -> pyruvate + D-  
CC glutamate.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U26947; AAB50428.1; -.  
DR HSSP: P19938; IDAA.  
DR InterPro: IPR001544; AminoTran\_4.  
DR Pfam: PF01063; AminoTran\_4; 1.  
DR ProDom: PD001961; AminoTran\_4; 1.  
DR TrGFams: TIGR01121; D.amino.amino; 1.  
DR PROSITE: PS00770; AA\_TRANSFERS\_CLASS\_4; 1.  
DR Transfaser: AminoTransferase; Pyridoxal phosphate.  
DR BINDING 144 144 PYRIDOXAL PHOSPHATE (POTENTIAL).  
DR SEQUENCE 283 AA; 31396 MW; 2CABFA604FEDE9D7 CRC64;  
SQ  
Query Match 17.5%; Score 309.5; DB 1; Length 283;  
Best Local Similarity 30.7%; Pred. No. 14e-17;  
Matches 85; Conservative 49; Mismatches 112; Indels 31; Gaps 7;  
QY 52 GEYLPAAEAKISIFDTGFHSDLYTYVAHWHGNIIFRLGDHRLDLGASKRLADAGYSK 111  
DB 7 GLMERSCAVDIEDRGVFGDGYEVIRIYNGILFTLDEHIAIKYSAEIGDLSFSE 66  
QY 112 DELAEITKCVSMQSLRESFVNLVYTGKGRKEKDKSLTHOVYIYALPYLMAFPPE 171  
DB 67 AELSSQLEKLVINDORRGGLYLVYTGKAPRKQYQ-AGLTPQVATYTPPI--QKPEKE 123  
QY 172 QIFGTATVPRHVRAGRNVDPTIKNYQ--GDLTAASF-----EAKDRGARPAIL 221  
DB 124 QONGVSAL-----TADMKMLKCDIKSLINLYNMVMIKQKQOASAFEPAIL 168  
QY 222 LQSDNCAVEGGEFVNCYIKDKLAS--PSRNALPGITRKVFEIADQMGIEATLADYTSR 279  
DB 169 I-RGLVTEGTSNMVYAKQNVITHTPYTLINGITRMVYLQCEENGLNVEKATYTKD 227  
QY 280 ELIYADELMAVYTAGVTPINSLDGEAVNGEPPPLT 316  
DB 228 ELINAEVEFITSTAEVITPYSIDGQTIGSGAPGPLT 264

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Branched-chain amino acid aminotransferase (EC 2.6.1.42) (BCAT).  
GN LIVE OR PA5013.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxId=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PAK.  
RA Franklund C.V., Coyne M.J., Goldberg J.B.;  
RT "Clustering of the lipopolysaccharide core genes, waaf, waac, waag,  
RT and waap, in Pseudomonas aeruginosa."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RA MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Salter M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RL Nature 406:959-964(2000).  
CC -1- FUNCTION: ACTS ON leucine, isoleucine and valine (By similarity).  
CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate -> 4-methyl-2-  
CC oxopentanoate + L-glutamate.  
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
CC AMINOTRANSFERASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: U63816; AAC33172.1; -.  
DR EMBL: AE004913; MAG08398.1; -.  
DR HSSP: P00510; IAC3G.  
DR InterPro: IPR001544; AminoTran\_4.  
DR Pfam: PF01063; AminoTran\_4; 1.  
DR ProDom: PD001961; AminoTran\_4; 1.  
DR TrGFams: TIGR01122; LIVE\_1; 1.  
DR PROSITE: PS00770; AA\_TRANSFERS\_CLASS\_4; 1.  
DR Transfaser: AminoTransferase; Branched-chain amino acid biosynthesis;  
DR Pyridoxal phosphate; Complete proteome.  
DR BINDING 160 160 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
DR CONFLICT 85 104 RAAVENNIDESATYTPMVFY -> APPCAATWAPISARW  
DR FT CST (IN REF. 1).  
DR FT A -> S (IN REF. 1).  
DR FT SEQUENCE 124 124 5CD5CC6CFCEDE423E CRC64;  
SQ  
Query Match 17.3%; Score 304.5; DB 1; Length 307;  
Best Local Similarity 30.0%; Pred. No. 3.9e-17;  
Matches 93; Conservative 47; Mismatches 117; Indels 53; Gaps 10;  
QY 46 GVAMIGETLPAEAKISIFDTGFHSDLYTYVAH--VWHG-----NIFRLDH 92  
DB 8 GVIWYDGLVQMRDATHY-----LTHLHVGMYFEGVRAVDTPQGTATRLDAH 58  
QY 93 LDRLLDGASKRLADAGYSQDELAETKCVSMQSLRESFVNLV---TRGYGRK-KGEKD 148  
DB 59 TDLRFDSAHIMNMQIPYSHDEINEATRAAVRENNEISAVIRMYVYGGSGGLRASGLK- 117  
QY 149 LSKLTHOVYIYALPYLMAFPPEAOLF-----GTLATVPRHVR--RAGRTVDPPTIKN 198

Db 118 -----VHVTIAMSNGVAGEALQGIKVTSTFRHHVNSMTRAKSNCA----- 164  
 QY 199 YOMGDLTAASFEAKDKARAIILDSNCAEGEPNCCVKGKGLASSRNA-LEGTIR 257  
 Db 165 --YINSMAIQAQEAISGADAEAMMIDPEGYVAGESSNIFIKGVIYTPPTACINGLIR 222  
 QY 258 KTYEADLQMGIEATLADVTSRELYDADELAATVGTGVPINSLDGEAVNGEPEPLTV 317  
 Db 223 NTITLTAEGFKLVEKRTIRDEVYLADEAFETGTAAEVTPIREVDGRKIGAGRRGPVE 282  
 QY 318 AIRDFWALM 327  
 Db 283 KLGKAYFDLV 292

RESULT 7  
 ID LIVE\_SALTY STANDARD; PRT; 308 AA.  
 AC P15168; 091670;  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Branched-chain amino acid aminotransferase (EC 2.6.1.42) (Transaminase B) (BCAT).  
 GN LIVE OR STM3903 OR STMD1.89 OR STY3654.  
 OS Salmonella typhimurium, and  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 CX NCBI\_TaxID=602, 601;  
 RX MEDLINE=9353621; PubMed=2669973;  
 RA Feld M., Nguyen D.C., Armstrong F.B.;  
 RT "Amino acid sequence of Salmonella typhimurium branched-chain amino acid aminotransferase."  
 RL Biochemistry 28:5306-5310(1989).  
 RN 121  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S typhimurium; STRAIN=LT2 / SCS1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanders K.E., Spleth J., Clifton S.W., Latreille P., Courtney L., Portwolk S., Ali J., Dante M., Du F., Hou S., Layman D., Leong S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium LT2."  
 RL Nature 413:852-856(2001).  
 RN 131  
 RP SEQUENCE OF 1-5.  
 RC SPECIES-S typhimurium;  
 RA Randall R.R., Wallis M.H., Young G.J., Armstrong F.B.;  
 RT "N-terminal sequence of branched-chain amino acid aminotransferase."  
 RL Fed. Proc. 38:325-325(1979).  
 RN 141  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-S typhi; STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebalina M., Baker S., Basham D., Brooks K., Chillingworth T., Connerton P., Cronin A., Davis P., Davies R.M., Dew L., White N., Farrar J., Feltwell T., Hamlin N., Hogue A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RT "Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine.  
 -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate -> 4-methyl-2-

CC oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
 CC -1- SUBUNIT: HOMOHETEROMER.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF233324; AAF33481.1; -  
 DR EMBL; AE008882; AAL22753.1; -  
 DR EMBL; AL627279; CAD09414.1; -  
 DR PIR; A05077; A05077.  
 DR PIR; A34082; A34082.  
 DR HSSP; P00510; 1A3G.  
 DR StyGene; SG10181; 1LVE.  
 DR InterPro; IPR01544; AminoTran\_4.  
 DR Pfam; PF01063; aminoTran\_4; 2.  
 DR ProDom; PD001961; AminoTran\_4; 1.  
 DR TIGRFAMs; TIGR01122; 1LVE; 1.  
 DR PROSITE; PS00770; AA\_TRANSFERR\_CLASS\_4; 1.  
 DR TRANSFAM; TIGR01122; 1LVE; 1.  
 KW Transferrase; Aminotransferase; Branched-chain amino acid biosynthesis;  
 KW Pyridoxal phosphate; Complete proteome.  
 FT INIT\_MET 0  
 FT BINDING 159  
 FT SEQUENCE 308 AA; 33921 MW; 8F826434F9386AAD CRC64;  
 SQ

Query Match 17.1%; Score 302.5; DB 1; Length 308;  
 Best local Similarity 31.6%; Pred. No. 5.6e-17;  
 Matches 91; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

QY 49 WIRGEYLPAPAEAKISIFDGFHSDLTYYVAHWGN-----IFRLGDHDLRLDGASKIR 104  
 Db 9 WFNQGVWREWDAAKAVHMSHALHGTSTVFEGIRCYDSKHPVFRHREHQRILDSAKYIR 68  
 QY 105 LDAGYKDELAEITTKKCVMSQLSRESFNVLTVRGKKRKGKEDLSKTHOYIYAIPLV 164  
 Db 69 FPKYSQSIDELMEKCRVIRKNNLTSAIRPLRVG-DVKGAVNPPPGYITVDYIAAFW- 126  
 QY 165 WAFPPAEQI-FGTTAIVPRVRRAGRNTVDPTIK--NTQMGDLTAASFEAKDKARAI 220  
 Db 127 GAVLGAELADQIDAMVSW-NRAAPNTIPTAKAGNVLSSLVGS-EARRHGQEG 183  
 QY 221 LLDSDNCVAGEPNCCVYVDGKLASP--SRNALPGITRTKTYPELADMGIEATLADVTS 278  
 Db 184 ALDVNGYISGAGENLFEVKGDLFTPPFTSSALPGITRTAIIKILAKEIGIEVREOVLSR 243  
 QY 279 RELYDADELAATVGTGVPINSLDGEAVNGEPEPLTVAIRDFWAL 326  
 Db 244 ESLYLADEVMSTGAETITPVRSVDQIGVEGNCGVYTRIKQAIFGL 291

RESULT 8  
 ID LIVE\_ECOLI STANDARD; PRT; 308 AA.  
 AC P00510; 047299;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Branched-chain amino acid aminotransferase (EC 2.6.1.42) (Transaminase B) (BCAT).  
 GN LIVE OR B3770 OR 25281 OR ECS4704.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CX NCBI\_TaxID=562, 83334;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE=85289113; PubMed=3897211;  
 RX Kuramitsu S., Ogawa T., Ogawa H., Kagamiyama H.;  
 RT "Branched-chain amino acid aminotransferase of *Escherichia coli*:  
 RL nucleotide sequence of the *lve* gene and the deduced amino acid  
 RN sequence.";  
 RP J. Biochem. 97:993-999(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE=87174741; PubMed=3550695;  
 RX Lawther R.P., Wek R.C., Lopes J.M., Pereira R., Tallon B.E.,  
 RA Hatfield G.W.;  
 RT "The complete nucleotide sequence of the *livGMDA* operon of  
 RL *Escherichia coli* K-12.";  
 RN Nucleic Acids Res. 15:2137-2155(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=92358234; PubMed=1379743;  
 RA Daniels D.L., Plunkett G., III, Burland V.D., Blattner F.R.;  
 RT "Analysis of the *Escherichia coli* genome: DNA sequence of the region  
 RL from 84.5 to 86.5 minutes.";  
 RN Science 257:771-778(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anahtarman T.S., Iln J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsuo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Ikida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,  
 RA Kunara S., Shibata T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 RL O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RN DNA Res. 8:11-22(2001).  
 RN [6]  
 RP PRELIMINARY SEQUENCE OF 1-80 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE=80101081; PubMed=392469;  
 RA Lawther R.P., Nichols B.P., Zurawski G., Hatfield G.W.;  
 RT "The nucleotide sequence preceding and including the beginning of the  
 RL *lve* gene of the *livGMDA* operon of *Escherichia coli* K12.";  
 RN Nucleic Acids Res. 7:2289-2301(1979).  
 RN [7]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RC MEDLINE=92235862; PubMed=1569580;  
 RX Pregel J.M., Winkelman J.W., Adams C.W., Hatfield G.W.;  
 RT "DNA topology-mediated regulation of transcription initiation from  
 RL the tandem promoters of the *livGMDA* operon of *Escherichia coli*.";  
 RN J. Mol. Biol. 224:919-935(1992).  
 RN [8]  
 RP PYRIDOXAL-PHOSPHATE BINDING SITE.  
 RX MEDLINE=89174510; PubMed=3069845;  
 RA Inoue K., Kuramitsu S., Aki K., Watanabe Y., Takagi T., Nishigai M.,  
 RA Ikai A., Kagamiyama H.;  
 RT "Branched-chain amino acid aminotransferase of *Escherichia coli*:  
 RL overproduction and properties.";  
 RN J. Biochem. 104:777-784(1988).  
 RN [9]

X-RAY CRYSTALLOGRAPHY (2.5-ANGSTROMS).  
MEDLINE:97306043; PubMed:9163511).  
Okada K., Hirotsu K., Sato M., Hyashi H., Kagamiyama H.:  
"Three-dimensional structure of Escherichia coli branched-chain amino  
acid aminotransferase at 2.5-A resolution.".  
J. Biochem. 121:637-641(1997).  
-1- FUNCTION: Acts on leucine, isoleucine and valine.  
-1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate -> 4-methyl-2-  
oxopentanoate + L-glutamate.  
-1- COFACTOR: PYRIDOXAL PHOSPHATE.  
-1- PATHWAY: Valine and isoleucine biosynthesis.  
-1- SUBUNIT: HOMOHExAMER.  
-1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
AMINOTRANSFERASES.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

[illegible]

	Query Match	16.9%	Score 298.5	DB 1,	Length 308;
	Best Local Similarity	30.9%	Pred. No. 1.2e-16;		
	Matches	89;	Conservative	49;	Mismatches 135; Indels 15; Gaps 8;
OY	49 WIEGEVLPAAEAKISIFDTFGCHSDLTYYVAHVHGN----	IPIRIGDHLRLDLGASKLR	104		
	: : : :   :	: : : : :   :			
Db	9 WENGEMRWMDAAKVHSMHALHYGTSTVEGRICDYDSHGKGPVFYFHRRHMGRLMDSAKIYR		68		
OY	105 LDNGYSMDLAEITKKCVSMQSOLRSPFNULTYTRGYCKRKKEKLDSLTHQVITYAIPLI		164		
	: : : : :   :	: : : : :   :			
Db	69 FPYOSQSIDELMECKRCRVIRKNNLTSAYIRPLIFVG-DVGMGVNRPAGYSDVIYAAPFM-		126		
OY	165 WAEPPEAQI-FGTATAVPRPHVRARRARIVNDPTIK---	NYOMGDLTAASFPAKRGARTAI	220		
	:   :   :   :   :	:   :   :			
Db	127 GALIGDALEOGGIDANVSWM-NRAAPNTIIIPAAKAGNVLSLLVGS--BARHHGOECI		183		
OY	221 LLDSDNCVAEGGFNVCIYKDGKLASP--SNNAIPGITRTKTFVELADOMGIEATLDDYTS		278		
	: : :   :   :   :   :   :	: : :   :   :   :   :   :			
Db	184 ALDYNVICISEGAGENLFYEKDSVLTEPFTSSALPGITROIILIKLNELCIEFREQVLSR		243		
OY	279 RELYDDELMAVTYAGCTVPINSLDGEAVNGDEGPLTVAIRDRPMAL		326		
	:   :   :   :   :   :	:   :   :   :   :   :   :			









Db 279 P 279

RESULT 12

DATA\_BACSU STANDARD: PRT: 282 AA.

ID DATA\_BACSU STANDARD: PRT: 282 AA.

AC 007597:

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate

DE aminotransferase) (D-amino acid aminotransferase) (D-amino acid

DE transaminase) (DAAAT).

GN DAT.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.,

RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,

RA Borris R., Bourcier L., Brans A., Braun M., Bridgell S.C., Bron S.,

RA Boullelet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,

RA Chol S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,

RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,

RA Frits C., Fujita Y., Fujita Y., Fuma S., Gallizzi A., Galleron N.,

RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,

RA Giusseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudega B., Park S.H.,

RA Parro V., Pohl T.M., Portetlelle D., Potwolik S., Prescott A.M.,

RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,

RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Zimstein A.,

RT "The complete genome sequence of the Gram-positive bacterium Bacillus

subtilis."

RL Nature 390:249-256(1997).

CC -1- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,

GLUTAMATE, AMINO BUTYRATE, NORVALINE AND ASPARAGINE (By

similarity).

CC -1- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate -> pyruvate + D-

glutamate.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT

AMINOTRANSFERASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation-

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: Y14082; CAA74512.1; -

DR EMBL: Y14080; CAA74451.1; -

DR EMBL: Z99109; CAB12806.1; -

DR HSSP: P19938; 2DAB.

DR Sublist: BG13045; dat.

DR InterPro: IPR001544; AminoTran\_4.

DR Pfam: PF01063; AminoTran\_4; 1.

DR ProDom: PD001961; AminoTran\_4; 1.

DR TIGRfams: TIGR01121; D-amino\_aminot; 1.

DR PROSITE: PS00770; AA\_TRANSFERP; CLASS.4; 1.

KW Transferase; Amino transferase; Pyridoxal phosphate; Complete proteome.

FT BINDING 144 144 PYRIDOXAL PHOSPHATE (POTENTIAL).

FT SEQUENCE 282 AA: 31182 MW: 495587565DFDECI CRC64;

Query Match 15.8%; Score 278.5; DB 1; Length 282;

Best Local Similarity 29.1%; Pred. No. 4; Le-15;

Matches 82; Conservative 48; Mismatches 129; Indels 23; Gaps 6;

OY 50 IGEYLPAAEAKISIFDTGFGSHDITTYVAHVHGNIFRLGDHLLDRLLDGASKRLDAGY 109

DB 5 VNGRLIGRSEASIDEDRGYOGDGIYEVIRYGVGLREHARFPRSAEIGISLDF 64

OY 110 SMDLEAITKKCVSMQSOLRESFVNLTVRGYGRKGEKLSKLTQVYIYVPLMAE-- 167

DB 65 STEDLEMDLQKLVQENAVSEGAIVYIQTTRGVAPR-----H0YEGALEPOTYATF 115

OY 168 ---PPAQDITGTVAVIRHVRAGRNTVPTIKYWGMDLTAASFKAQDCAFTAILD 223

DB 116 TYKREQDQNAVVAITDEDLRWL---RCDIKSLNLTYNWTKQ--RAYEGAFEAILL- 169

OY 224 SDNCVAEPPGVNCIVKDGKLA--PSRNALPGITRTKTVFELADQMGIATLVTSREL 281

DB 170 RGVVTEETSSNVYAVINGVTRHPANRLINGITRMNILLIKRNGKIKIDETVSEEL 229

OY 282 YDADLMVATAGVTPINSIDGEAVNGEPPGLTVALRDF 323

DB 230 KQAEFTISSTAEITPVTLIDGOSTGSKPGKPYTKOLAAE 271

RESULT 13

DATA\_LISTIN STANDARD: PRT: 289 AA.

AC 092890:

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate

DE aminotransferase) (D-amino acid aminotransferase) (D-amino acid

DE transaminase) (DAAAT).

GN DAT OR DATA OR LIN160.

OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OX NCBI\_TaxID=1642;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CLIP 11262 / Serovar 6a;

RX MEDLINE=21537279; PubMed=11679669;

RA Glaser P., Frangeul L., Buchrieser C., Rusnick C., Amend A.,

RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Darvar A., Denoux F.,

RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,

RA Entian K.-D., Fshl H., Garcia-del Portillo F., Garrido P.,

RA Gantier L., Goebel W., Gomez-Lopez N., Hahn T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Kreft J., Kohn M., Kunst F., Kunaprat G.,

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schlueber T., Simoes N., Tierrez A.,

RA Vaquer-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of *Listeria species*."

RL Science 294:849-852(2001).

CC -1- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,

GLUTAMATE, AMINO BUTYRATE, NORVALINE AND ASPARAGINE (By

CC similarity).

CC -1- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-

CC glutamate.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT

CC AMINOTRANSFERASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL: AL596169; CAC96891.1; -

CC DR Listlist: L1N01660; -

CC DR InterPro: IPR001544; AminoTran\_4.

CC DR Pfam: PF01063; AminoTran\_4; 1.

CC DR ProDom: PD001961; AminoTran\_4; 1.

CC DR TIGRFAMs: TIGR01121; D.amino.aminoT; 1.

CC DR PROSITE: PS00770; AA\_TRANSFERRASE; Pyridoxal phosphate (POTENTIAL).

CC KW Transferrase; AminoTranferase; Pyridoxal phosphate; Complete proteome.

CC FT BINDING 147 147 PYRIDOXAL PHOSPHATE (POTENTIAL).

CC SQ SEQUENCE 289 AA: 3323 MW: 30805951384173 CRC64:

Query Match 15.0%; Score 264; DB 1; Length 289;

Best Local Similarity 28.1%; Pred. No. 6.1e-14;

Matches 78; Conservative 51; Mismatches 119; Indels 30; Gaps 6;

OY 59 EAKISIEPTGFGHSDLYTVAHVHGNIFRLGDLRLDLDGASKRLDAGYSKDELAET 118

DB 14 DATVVEEDRGYFGDGYEVRLYNGKFFTYNEHIDRLYASAKRIDVLPYSKETLALL 73

OY 119 KCVMSQLRSEFVLTPTRGYGRKRGKDKSLTHQYIYAIYMAFPADQIFGT 178

DB 74 DKLVAENINNTGNVLTQYTRGVQNDPRN-----HVLDP--DFPEGLVTAAR 118

OY 179 IVPFRHVRAGRNVTDPITKNYQW-----GDILTAASFEAKDKGAFATLIDSDNC 227

DB 119 EYPRNERQPIEGGSAITBEDVRLMKDLSKLNGLNLAKN-KAHQOALALAILHRGNO- 176

OY 228 VAEQGEFVNCIVKDGKLAASPSRN--ALPGITRKTYVELADMGIEATLRDYSSELYDAD 285

DB 177 VTCSASNVSIKDGVLMTTHADNLMLNGITROVLIIDVAKKNGIPVKEADPTLIDLRDAD 236

OY 286 ELMAVTAGVYPIINSIDGVAENGEPGLVATIRDRF 323

DB 237 EYFISSTIETPTIHDGVYADGKRPITRQJLHNYF 274

RESULT 14

DATA\_BACSH STANDARD: PRT: 283 AA.

AC P54693; 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE D-alanine aminotransferase (EC 2.6.1.21) (D-aspartate

DE aminotransferase) (D-amino acid aminotransferase) (D-amino acid

DE transaminase) (DAA).

GN DAT.

OS Bacillus sphaericus.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

CC NCBI\_TaxID=1421;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 10208;

RA Fotheringham I.G., Taylor P.P., Bledig S.A.;

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 28;

RA Hwang T.S., Wang N.C., Wu S.P., Yang C.L., Tsai H.;

RT Cloning and overexpression of D-amino acid aminotransferase from

RT Bacillus sphaericus in *Escherichia coli*;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ACTS ON THE D-ISOMERS OF ALANINE, LEUCINE, ASPARTATE,

CC GLUTAMATE, AMINOBUTYRATE, NORVALINE AND ASPARAGINE (By

CC similarity).

CC -1- CATALYTIC ACTIVITY: D-alanine + 2-oxoglutarate = pyruvate + D-

CC glutamate.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.

CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT

CC AMINOTRANSFERASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC DR EMBL: U26732; AAA68028.1; -

CC DR EMBL: AF081278; AAC33964.1; -

CC DR HSSP: P19938; IDAA.

CC DR InterPro: IPR001544; AminoTran\_4.

CC DR Pfam: PF01063; AminoTran\_4; 1.

CC DR ProDom: PD001961; AminoTran\_4; 1.

CC DR TIGRFAMs: TIGR01121; D.amino.aminoT; 1.

CC DR PROSITE: PS00770; AA\_TRANSFERRASE; Pyridoxal phosphate (POTENTIAL).

CC KW Transferrase; AminoTranferase; Pyridoxal phosphate.

CC FT BINDING 146 146 PYRIDOXAL PHOSPHATE (POTENTIAL).

CC SQ SEQUENCE 283 AA: 31477 MW: 67AD497D5CA01ABD CRC64:

Query Match 14.2%; Score 250; DB 1; Length 283;

Best Local Similarity 26.9%; Pred. No. 7.7e-13;

Matches 79; Conservative 49; Mismatches 110; Indels 56; Gaps 9;

OY 58 EAKISIEF-DTGFHSDLYTVAHVHGNIFRLGDLRLDLDGASKRLDAGYSKDELA 115

DB 12 EESGISTSPEDRGYFGDGYEVRLYNGKFFTYNEHIDRLYASAKRIDVLPYSKEDV 71

OY 116 EITKCVMSQLSEFVLTPTRGYGRKRGKDKSLTHQYIYAIYMAFPADQIFGT 175

DB 72 KLLHDLLEKNNLMTGVHYPQITRGTSRN-----HIPPDA-----S 107

OY 176 TPAIVPRHVRAGRNVTDPITKNYQW-----DLTA-----ASPEARD 213

DB 108 VPAVLIGNVK-----TGENSIENFEKGVKATIVEDVRLMKDLSKLNGLAVLAQGEASE 162

OY 214 RGARTAILDSDNCVAEGPEFVNCIVKDKL-ASPSRN-ALPGITRKTYVELADMGIEA 271

DB 163 KCGYEAILHRGD-IITECSANVYGKIDGKLYHPANNYILNGITROVLIKCAAEINLPV 221

OY 272 TLADVYSRELYDDELAMNTAGVYPIINSIDGVAENGEPGLVATIRDRFWA 325

DB 222 IEEPMTKGLDILMDELIVSSSEVTPVLDVQDQIGAGVPDEMTKLOKAFEA 275

RESULT 15

LIVE\_RICCN STANDARD: PRT: 290 AA.

AC Q92126; 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable branched-chain amino acid aminotransferase (EC 2.6.1.42)

DE (BCAT).

GN LIVE OR RC0594.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI\_TaxID=781;

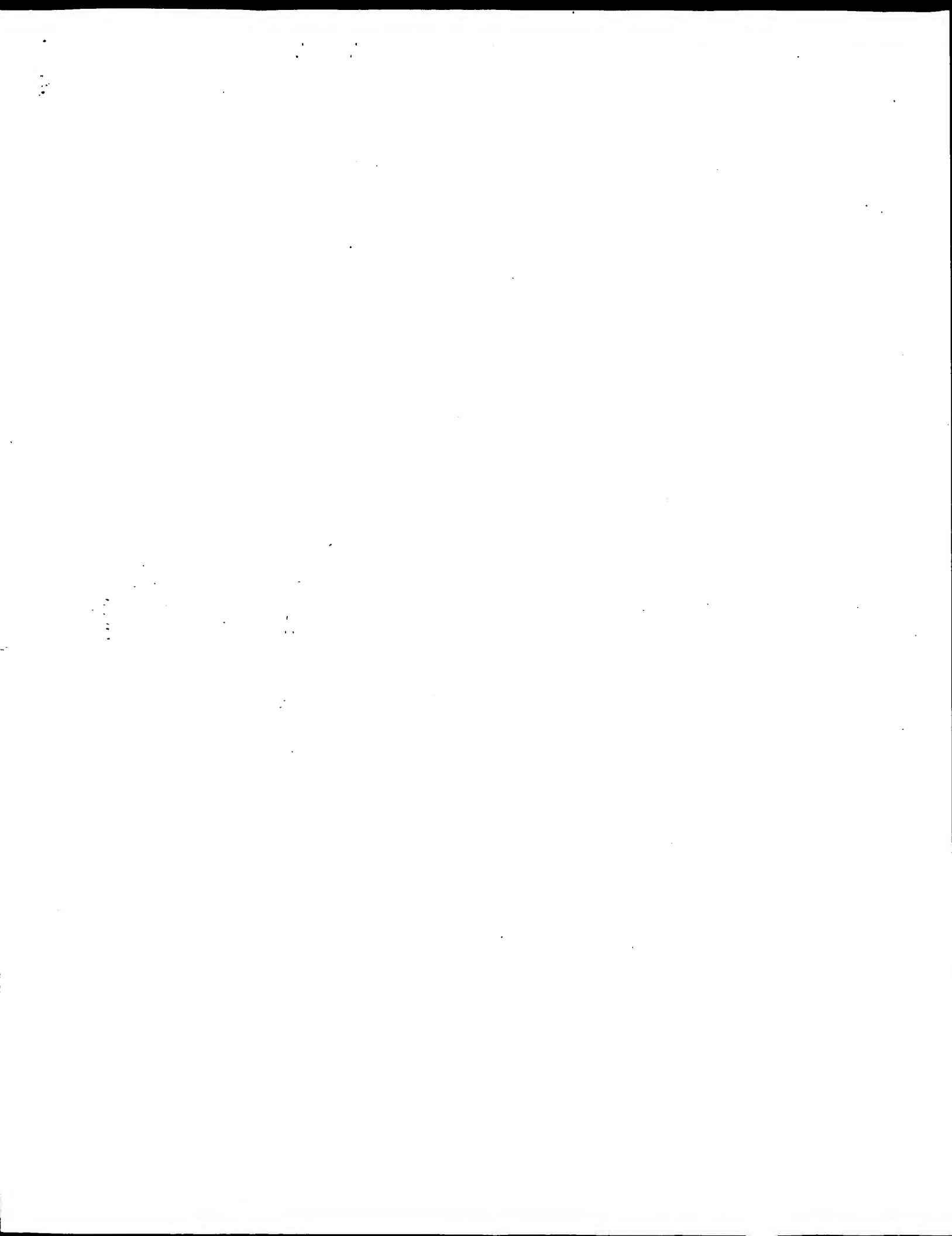
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-Malish 7;  
 RX MEDLINE=21442074; PubMed=11557893;  
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,  
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,  
 RA Raoult D.;  
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";   
 RL Science 293:2093-2098(2001).  
 CC -1- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).  
 CC -1- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-  
 CC oxopentanoate + L-glutamate.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: Valine and isoleucine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT  
 CC AMINOTRANSFERASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE008620; AAL03132.1; -  
 DR InterPro: IPR001344; AminoTran\_4.  
 DR Pfam: PF01063; AminoTran\_4; 1.  
 DR ProDom: PD001961; AminoTran\_4; 1.  
 DR PROSITE: PS00770; AA\_TRANSFER\_CLASS\_4; 1.  
 DR Transferrase; AminoTransferase; Branched-chain amino acid biosynthesis;  
 KM Pyridoxal phosphate; Complete proteome.  
 FT BINDING 155 155 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 SQ SEQUENCE 290 AA; 33243 MW; 86112A254A505177 CRC64;

Query Match 12.0%; Score 212; DB 1; Length 290;  
 Best local similarity 24.7%; Pred. No. 8.6e-10;  
 Matches 67; Conservative 52; Mismatches 124; Indels 28; Gaps 5;

QY 49 WIEGEYLPAEAKISIFDTGFGHSDLTYYVAHWGNIERLGDHDLRLDGASKRLDAG 108  
 || :  
 DB 14 WINGDLVYPQFARLHVTHSLHSGSVFEGERAINGKVFELKEHTARLTKSAAELGLKVP 73  
 || :  
 QY 109 YSKDELAETIKKCVSMQLESFV-----NLVTGKGKRGKSKLTHOYIY 159  
 || :  
 DB 74 YNVEDEIIKAHECVYKONNINDAYIRPLIMCGDSESLNTNOY-----LSTNLLIA 122  
 || :  
 QY 160 AIPPLMAFPPEAQIFGTTALVPRHVRAGRTVDPTIKNYOGDGLTAASPEAKDRGART 219  
 || :  
 DB 123 GTPSM-----PRSEFGKINLHVSRRKAMPSTPYQSKSAQIMAITSKKAKALGYEDA 178  
 || :  
 QY 220 ILDSDNCAVEGPGFNVCYKDGKLSP--SRNALPGITRTKVPFELADOMGIEATLRDVT 278  
 || :  
 DB 179 LLDDYEGYIAECTTNTNFYFKDKILYPIADRFPLNGITROTITEIAKDLGLEVKERLKL 238  
 || :  
 QY 279 RELYDADELMAVTTAGGYTPINSIDGEAVGN 309  
 || :  
 DB 239 EOIEDFTGCEVTGTAIEVQNMIDID--LGN 266

Search completed: April 17, 2003, 17:37:55  
 Job time : 27 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:35:22 ; Search time 88 Seconds

(without alignments)  
793.750 Million cell updates/sec

Title: US-10-067-291-1

Perfect score: 1764  
Sequence: 1 MIALSDLGTSNLVAVEPGAI.....RDRFWALNDPEGLIETLEY 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_21:\*  
1: sp\_Archaea:\*  
2: sp\_Bacteria:\*  
3: sp\_Fungi:\*  
4: sp\_Human:\*  
5: sp\_Invertebrate:\*  
6: sp\_Mammal:\*  
7: sp\_Mhc:\*  
8: sp\_Organelle:\*  
9: sp\_Phage:\*  
10: sp\_Plant:\*  
11: sp\_Rodent:\*  
12: sp\_Virus:\*  
13: sp\_Vertebrate:\*  
14: sp\_Unclassified:\*  
15: sp\_Virus:\*  
16: sp\_Bacteriap:\*  
17: sp\_Archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	596.5	33.8	320	16	Q98K82 rhizobium 1
2	577	32.7	291	16	Q981W8 rhizobium 1
3	414.5	23.5	295	17	Q81UX5 methanopyru
4	394	22.3	294	17	Q8108 methanosa
5	352.5	20.0	291	16	Q8RB89 thermosar
6	340.5	19.3	553	10	Q8W027 arabidopsi
7	326.5	18.5	559	10	Q9ASR4 arabidopsi
8	318.5	18.1	490	10	Q9MAA0 arabidopsi
9	315	17.9	303	17	Q8ZRE2 pyrobacul
10	301	17.1	308	16	Q8ZAB2 versinia pe
11	298	16.9	308	17	Q9HNF8 halobacteri
12	291	16.5	282	2	Q9KW26 staphylococ
13	290	16.4	282	16	Q99TB4 staphylococ
14	289.5	16.4	290	16	Q8XRW7 raietonia s
15	278	15.8	287	16	Q92G85 rhizobium m
16	277.5	15.7	319	16	Q9KVV9 vibrio chol

17	273.5	15.5	296	16	Q9A4A8	Q9A4A8 caulobacter
18	273	15.5	293	16	Q8YD14	Q8YD14 brucella me
19	265.5	15.1	309	16	Q81X10	Q81X10 raietonia s
20	263	14.9	289	2	Q937U0	Q937U0 brucella su
21	260	14.7	283	16	Q9K939	Q9K939 bacillus ha
22	259	14.7	319	16	Q8UC16	Q8UC16 agrobacteri
23	256	14.5	287	16	Q98MX3	Q98MX3 rhizobium 1
24	253	14.3	304	16	Q9PTM6	Q9PTM6 campylobact
25	247.5	14.0	297	16	Q98HX8	Q98HX8 rhizobium 1
26	246	13.9	290	16	Q8UJK4	Q8UJK4 agrobacteri
27	244.5	13.9	285	17	Q9H564	Q9H564 halobacteri
28	240	13.6	287	16	Q9A7H9	Q9A7H9 caulobacter
29	238.5	13.5	293	16	Q930Z7	Q930Z7 rhizobium m
30	223.5	12.7	296	10	Q9FDT9	Q9FDT9 arabidopsi
31	217	12.3	290	2	Q9AKK0	Q9AKK0 rickettsia
32	212.5	12.0	295	10	Q9FTU3	Q9FTU3 oryza sativ
33	208	11.8	290	2	Q9AKQ4	Q9AKQ4 rickettsia
34	205.5	11.6	280	16	Q97KK4	Q97KK4 clostridium
35	204.5	11.6	271	16	Q9HZN6	Q9HZN6 pseudomonas
36	199	11.3	303	2	Q93H01	Q93H01 streptomyce
37	197.5	11.2	288	2	Q9AKES	Q9AKES rickettsia
38	178.5	10.1	359	2	Q9F329	Q9F329 staphylococ
39	176.5	10.0	273	16	Q9LID6	Q9LID6 streptomyce
40	172.5	9.8	292	16	Q98SK4	Q98SK4 rhizobium 1
41	165.5	9.4	217	2	Q9AM72	Q9AM72 rhodospirill
42	164.5	9.3	410	10	Q9M220	Q9M220 arabidopsi
43	161.5	9.2	418	10	Q9SNY9	Q9SNY9 solanum tub
44	161	9.1	276	2	Q9F7N1	Q9F7N1 uncultured
45	160.5	9.1	295	16	Q9KGH0	Q9KGH0 bacillus ha

## ALIGNMENTS

RESULT 1  
Q98K82 PRELIMINARY: PRT: 320 AA.  
AC Q98K82; 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE Branched-chain amino acid transferase.  
GN MRA1594.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_Taxid=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MAFF303099;  
RX MEDLINE-21082930; PubMed-11214968;  
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT Complete genome structure of the nitrogen-fixing symbiotic bacterium  
Mesorhizobium loti.  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP002997; BAB4832.1; -  
DR InterPro: IPR001544; Aminostran\_4.  
DR Pfam: PF01063; aminostran\_4; 1.  
DR ProDom: PD001961; Aminostran\_4; 1.  
KW Transferase; Complete proteome.  
SQ SEQUENCE 320 AA: 34742 MW: 837EAFACACAGF614 CRC64;

Query Match 33.8%; Score 596.5; DB 16; Length 320;  
Best Local Similarity 40.7%; Pred. No. 3.6e-40;  
Matches 121; Conservative 57; Mismatches 116; Indels 3; Gaps 3;  
QY 34 DYELDTSPPAGVAMIEGFIPEAEAKISIFDTGSHDLTYTAHVWNGNIFRLGDLH 93  
DB 18 DRHVDPHS-VPDGIAFLDGLPMSQAKSVLDWMGFLHSDATYDTGVHWNNGRFRDLHL 76

QY 94 DRLDGLASRLRLDAGYSKDELAETKKCVSMQSLRESFVNLTFRGYKRRGENDLSKLT 153  
 DB 77 DREFGLEKRLMTIPEDRGVAEILHNCVALSGHRAAYEMLCSTRG-ASPTFSRDPQAI 135  
 QY 154 HGVYIATPYLMAFPAPDIFCTTAIVPRHVRACRNTVDPIKRYOMGDLTAASFEEKD 213  
 DB 136 NRPMAFAVFP-GSVANAEOQLRGCLRNAISDKYRIPASVDSIKNYHMDLVRLGLYDAYD 194  
 QY 214 RGARTAILDSDNCVAEGFENVCIVKDKLASPSRNALPGITKRTVEFLADOMGIEATL 273  
 DB 195 SGAEETAILDFNGNVAEGGFENVCIVKDKLASPSRNALPGITKRTVEFLADOMGIEATL 254  
 QY 274 RYVTSRELYADDELMAVTTAGVTPINSIDGFAVNGEPGRLTYAIRDRFALMDEP 330  
 DB 235 ADVSVAALKAADDEVITSTAGIMPEIDGMAIDAGKVGCVTSRLMALYQKHDP 311

## RESULT 2

ID Q981W8 PRELIMINARY; PRT; 291 AA.  
 AC Q981W8;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, last annotation update)  
 DE Branched-chain amino acid aminotransferase.  
 GN ML9205.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303039;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneke T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti";  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL, AP003015; BAB54591.1;  
 DR InterPro: IPR001544; AminoLtran\_4.  
 DR Pfam: PF01063; AminoLtran\_4; 1.  
 DR ProDom: PD001961; AminoLtran\_4; 1.  
 DR Transferrase; AminoLtran: plasmid; Complete proteome.  
 KW SEQUENCE 291 AA; 32309 MW; E7981CDLEABR3EFL CRC64;

Query Match 32.7%; Score 577; DB 16; Length 291;  
 Best Local Similarity 39.4%; Pred. No. 1.2e-38;  
 Matches 112; Conservative 60; Mismatches 110; Indels 2; Gaps 2;

QY 47 VAMTBGELPLPEAKISIFDTGFGHSDLTYYAAVHMGNTFRCLDHLRLDGLASRLRLD 106  
 DB 1 MAFMDGQYLPSEAKVSLDWMGFLHSDAIDYDVTVAHWGDRFRLMLHADRFRGMEKRLMK 60  
 QY 107 AGYSKDELAETTKKCVSMQSLRESFVNLTFRGYKRRGENDLSKLTNOYIYVPLMA 166  
 DB 61 LPRNSELKELSLTCAVLSGHSKSAFVETICRG-GSPTFSRDPQSNRRTAFAVFP-GS 118  
 QY 167 FPPAEQJFGTAAIVPRHVRACRNTVDPIKRYOMGDLTAASFEEKDARGARTAILDSDN 226  
 DB 119 VANKEQJERGLHVAISMVTRIPKSIDPTIKNYHMDLVGLFDAYGAFETALYIND 178  
 QY 227 CVASGPGFNCIVKDKLASPSRNALPGITKRTVEFLADOMGIEATLRYDTSRELYDADE 286  
 DB 179 NIAGPPEFNTVYDGLKLPAYGVLAGITRQTFEDLCDELGLSVASGIDIRRELKGADE 238  
 QY 287 LMAVTTAGVTPINSIDGFAVNGEPGRLTYAIRDRFALMDEP 330

DB 239 VFITSTAGIMPVSKIDETVVGDKVGLATRLROLADLYWEKHA DP 282

## RESULT 3

ID Q8TU5 PRELIMINARY; PRT; 295 AA.  
 AC Q8TU5;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Branched-chain amino acid aminotransferase.  
 GN IIVE OR MK1627.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Shesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
 RA Sheerbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Nataraj D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malyn A.G., Koonin E.V., Kozlovskiy S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 and monophyly of archaeal methanogens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL: AE010451; AAM02840.1;  
 KW Transferrase; AminoLtran: Complete proteome.  
 SQ SEQUENCE 295 AA; 32955 MW; 5DB962E95D549BB CRC64;

Query Match 23.5%; Score 414.5; DB 17; Length 295;  
 Best Local Similarity 33.0%; Pred. No. 1.8e-25;  
 Matches 97; Conservative 60; Mismatches 122; Indels 15; Gaps 4;

QY 49 WIGCEPLPEAKISIFDTGFGHSDLTYYAAVHMGNTFRCLDHLRLDGLASRLRLD 108  
 DB 9 YLNGEVLPRREAKISYVDFGLGDFVFGIRAYDRIFKLEBHDRLYDSAKAIMEIP 68  
 QY 109 YSKDELAETTKKCVSMQSLRESFVNLTFRGYKRRGENDLSKLTNOYIYVPLMAFP 168  
 DB 69 MTEKKREKALIEVVRANLDAIRVYVSRGEGDGLDPEKCEPVPVIT-----AE 120  
 QY 169 PABQIFCT-----TAIVPRHVRACRNTVDPIKRYOMGDLTAASFEEKDARGARTAILD 223  
 DB 121 PMBPLGDIYKKEIEVITASVRIPDADLPKIKSCNVLNILAKIQANLAGADEIMD 180  
 QY 224 SDNCVAEGGFENVCIVKDKLASPSRNA-LPGITKRTVEFLADOMGIEATLRYDTSRELY 282  
 DB 181 HEGYVCEGTGDNVFEVDETYTPEDTLIRGITRATVMEICEIGIPVEERITLGEIY 240  
 QY 283 DADELMAVTTAGVTPINSIDGFAVNGEPGRLTYAIRDRFALMDEP-PLIE 335  
 DB 241 AADEVLTGTAFAVAPVRAVDKRGKIGECPCGPIITRIMEAFRELTKKSTPYE 294

## RESULT 4

ID Q8TU5 PRELIMINARY; PRT; 294 AA.  
 AC Q8TU5;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Branched-chain amino acid aminotransferase.  
 GN IIVE OR M4349.  
 OS Methanosarcina acetivorans.  
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2214;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;  
 RX MEDLINE=21929760; PubMed=11932238;

Query Match	20.0%;	Score 352.5;	DB 16;	Length 291;
Best Local Similarity	31.2%;	Pred. No. 1.8e-20;		
Matches	91;	Conservative 55;	Mismatches 117;	Indels 29;
				Gaps 5.

	Query Match	Best Local	Similarity	Score	DB	Length	555;
	Matches	88;	Conservative	59;	Mismatches	138;	Indels 13; Gaps 5
QY	47	VAMTEGEEYLAEAFKASIPDTGFSDLTYTVAHVHNGNIFRGLDLRLDGAASKRLUD	106				
	:::		:		::		

```

Db 263 LAMWDELLPRMAKVSFDSVVGSGDSWEGRLTYKGIKFLREHDLRLDSAKALAFD 322
OY 107 AGYSKDELAEL-ITKCVMSQSLRESFVNLFTVRYGKRRKGEKDLSTLTHOYIYALPYL- 164
Db 323 NVPAEEVEALFRLTLNGMEDNTHIRLSLR-----GKKVTSQSPAINRRCCTLIIV 376
OY 165 ---WAFPPRQJFTGTAIVPRHVRAGNTVDPITKNTQMGDLTAASFEAKDGRARTAIL 221
Db 377 LAEMW-PPVYDNDGGIVLVATATRRNSPNNLDSKIHNNLNLNLAKIESNNNAADAIM 435
OY 222 LQSDNCVAGSGFNVCIYDKGLASPSRN-ALPGITRKVPELADOMGIEATLDDVTSRE 280
Db 436 LDKQSVSEINATINFMKKCVLTPHADYCLPGITRATVVELVYKENTLEERRISLSE 495
OY 261 LYDADELMAVTTAGVTPINSLDGEAVNGSGPPLTVAINRFRALMDEPGLIETIE 338
Db 496 FHTANEVTTGTMGELSPVVKIDGRVIGDGKGVPTRTLOAKYKKLLEDSCVPIPTYO 553

```

## RESULT 7

```

ID 09ASR4 PRELIMINARY; PRT; 559 AA.
AC 09ASR4;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ATG27410/F21A20_120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Chew R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carninci P., Chung M.K., Goldsmith A.D., Hayashizaki Y.,
RA Ithida J., Jones T., Kamiya A., Karlio-Neumann G., Kawai J., Lam B.,
RA Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Pham P.K., Quach H.L., Sakano H., Sakurai T., Satou M.,
RA Seki M., Southwick A., Toriumi M., Yamada K., Yu G., Shinozaki K.,
RA Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF367323; to the EMBL/GenBank/DBJ databases.
DR HSSP; P00510; 1A3G.
DR InterPro; IPR001544; Aminoctran_4.
DR Pfam; PF01063; aminoctran_4; 1.
DR ProDom; PD001961; Aminoctran_4; 1.
SQ SOURCE 559 AA; 62896 MW; 6A54870333F54793 CRC64;

```

Query Match 18.5%; Score 326.5; DB 10; Length 559;  
 Best Local Similarity 28.5%; Pred. No. 6.1e-18;  
 Matches 85; Conservative 59; Mismatches 141; Indels 13; Gaps 5;

```

OY 47 VANIEGEYLPABEAKISIFDTGFHSDLYTVVAHWGNIFRLGDLRLDLSGASRLD 106
Db 260 LAMWGDDEIVREMAKVSFDSVVGSGDSWEGRLTYKGIKFLREHDLRLDSAKALAFD 319
OY 107 AGYSKDELAEL-ITKCVMSQSLRESFVNLFTVRYGKRRKGEKDLSTLTHOYIYALPYL- 164
Db 320 NVPAEEVEALFRLTLNGMEDNTHIRLSLR-----GKKVTSQSPAINRRCCTLIIV 373
OY 165 ---WAFPPRQJFTGTAIVPRHVRAGNTVDPITKNTQMGDLTAASFEAKDGRARTAIL 221
Db 374 LAEMW-PPVYDNDGGIVLVATATRRNSPNNLDSKIHNNLNLNLAKIESNNNAADAIM 432
OY 222 LQSDNCVAGSGFNVCIYDKGLASPSRN-ALPGITRKVPELADOMGIEATLDDVTSRE 280
Db 433 LDKQSVSEINATINFMKKCVLTPHADYCLPGITRATVVELVYKENTLEERRISLSE 492
OY 261 LYDADELMAVTTAGVTPINSLDGEAVNGSGPPLTVAINRFRALMDEPGLIETIE 338
Db 493 FHTANEVTTGTMGELSPVVKIDGRVIGDGKGVPTRTLOAKYKKLLEDSCVPIPTYO 550

```

## RESULT 8

```

ID 09MAA0 PRELIMINARY; PRT; 490 AA.
AC 09MAA0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative branched-chain amino acid aminotransferase.
GN T12H1.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Mu D.,
RA Renning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T12H1 genomic sequence."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC009177; AAF27025.1;
DR HSSP; P00510; 1A3G.
DR InterPro; IPR001544; Aminoctran_4.
DR Pfam; PF01063; aminoctran_4; 1.
DR ProDom; PD001961; Aminoctran_4; 1.
KW Aminotransferase; Transferase.
SQ SOURCE 490 AA; 54829 MW; D29B6878564B36BD CRC64;

```

Query Match 18.1%; Score 318.5; DB 10; Length 490;  
 Best Local Similarity 28.0%; Pred. No. 2.2e-17;  
 Matches 82; Conservative 51; Mismatches 123; Indels 37; Gaps 3;

```

OY 47 VANIEGEYLPABEAKISIFDTGFHSDLYTVVAHWGNIFRLGDLRLDLSGASRLD 106
Db 232 LAMWGDDEIVREMAKVSFDSVVGSGDSWEGRLTYKGIKFLREHDLRLDSAKALAFD 291
OY 107 AGYSKDELAELITKCVMSQSLRESFVNLFTVRYGKRRKGEKDLSTLTHOYIYALPYL 166
Db 292 NVPAEEVEALFRLTLNVLAEL-----WK 316
OY 167 PEPADQIFGTGTAIVPRHVRAGNTVDPITKNTQMGDLTAASFEAKDGRARTAILDSN 226
Db 317 -PPVYDNDGGIVLVATATRRNSPNNLDSKIHNNLNLNLAKIESNNNAADAIMLDKG 375
OY 227 CVABEGPQVNCIVKDKGLASPSRN-ALPGITRKVPELADOMGIEATLDDVTSRELYDAD 285
Db 376 VYSEINATINFMKKCVLTPHADYCLPGITRATVVELVYKENTLEERRISLSEFTAN 435
OY 266 ELMAVTTAGVTPINSLDGEAVNGSGPPLTVAINRFRALMDEPGLIETIE 338
Db 436 EYWTGTGELSPVVKIDGRVIGDGKGVPTRTLOAKYKKLLEDSCVPIPTYO 488

```

## RESULT 9

```

ID 08ZTE2 PRELIMINARY; PRT; 303 AA.
AC 08ZTE2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Branched-chain amino acid aminotransferase (liver).
GN PAR3297.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;

```



RX PubMed-11792869;  
 RA Fitz-Gibbon S.F., Ladhner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.;  
 RT "genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL; AE009923; AAL64820.1;  
 DR Interpro; IPR001544; AminoTran\_4.  
 DR Interpro; IPR000577; FGGY\_kin.  
 DR Pfam; PF01063; aminoTran\_4; 1.  
 DR ProDom; PD001961; AminoTran\_4; 1.  
 DR TIGRfams; TIGR01122; lIve\_I; 1.  
 DR PROSITE; PS00770; AA\_TRANSFER\_CLASS\_4; 1.  
 DR PROSITE; PS00445; FGGY\_KINASES\_2; UNKNOWN\_1.  
 DR Transferrase; Complete proteome.  
 KM Transferrase; AminoTran; Complete proteome.  
 SQ SEQUENCE 303 AA; 34108 MW; EB518E6D1BDECA3 CRC64;

Query Match 17.9%; Score 315; DB 17; Length 303;  
 Best Local Similarity 28.0%; Pred. No. 2.1e-17;  
 Matches 85; Conservative 62; Mismatches 113; Indels 44; Gaps 11;

QY 42 PFAGVAMIGGEYLPAAEKISIFDTGFHSDLTITVAHVHGN--IFRLGDHLDRLL 97  
 DB 3 PYAKYI-WDGRILKWDKADKIHVTHAHGTSIFEGIRGYNMGDNLVPRLEHIDRMK 61  
 QY 98 DGASKRLDAGYSKDELAETKCKVMSQLE-----SFV-NLVTGCGYGRKGEKDL 149  
 DB 62 RSATILGINIPYREERYQAVLETKANFRREYIIRPAVFAVSQVY-----LDI 112  
 QY 150 SKLHGYIYAIPYLPAPPAEDIFGTTAVPRHVR-----RAGRVTDPITKNY--W 201  
 DB 113 RNLEVSIAIVFP-----FG-KYLSPMGKATIVSMRVRHATMLPMAKIGCIT 160  
 QY 202 GDLTAASEAKRDGARTAILDSDNCVAEGPFWNCIVKDKLASP--SHNALPGITRTK 259  
 DB 161 VNSTLALVEARSGFDALLMDVNGYVEGSGEIFIVRGRLFTPPVHSILLEGITROT 220  
 QY 260 VEEIADMGIEATLRDVTRELYDADELMAVTTAGVTPINSIDGAVNGEPGLTVAI 319  
 DB 221 VIKISGVNGLRVEERPYTREYVYADVFLVGTAEITPVEVDGRTIGCKPGPIITTKI 280  
 QY 320 RDRF 323  
 DB 281 AELY 284

RESULT 10

Q8ZAB2 PRELIMINARY; PRT; 308 AA.

AC Q8ZAB2;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Branched-chain amino acid aminotransferase (EC 2.6.1.42).  
 GN lIve OR YPO3899.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OC NCB1\_TaxID-632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;  
 RC MEDLINE-21470413; PubMed-11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Harraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Felwell T., Hamlin N., Holroyd S., Jagels K., Kariyasev A.V.,  
 RA Leather S., Moutle S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,  
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
 RL Nature 413:523-527(2001).  
 EMBL; AJ414159; CAC93365.1; -.

DR Interpro; IPR001544; AminoTran\_4.  
 DR Pfam; PF01063; aminoTran\_4; 1.  
 DR ProDom; PD001961; AminoTran\_4; 1.  
 DR TIGRfams; TIGR01122; lIve\_I; 1.  
 DR PROSITE; PS00770; AA\_TRANSFER\_CLASS\_4; 1.  
 DR Transferrase; AminoTran; Complete proteome.  
 KM Transferrase; Complete proteome.  
 SQ SEQUENCE 308 AA; 33864 MW; 574CB8C2B663380 CRC64;

Query Match 17.1%; Score 301; DB 16; Length 308;  
 Best Local Similarity 30.2%; Pred. No. 2.9e-16;  
 Matches 92; Conservative 46; Mismatches 141; Indels 26; Gaps 9;

QY 49 WIEGELPAEAKISIFDTGFHSDLTITVAHVHGN--IFRLGDHLDRLLDRLL 104  
 DB 9 WNGEAVPAEAKVYHSHALHYGTSVEGVYCSHGGPVYFRHREHRLHDSAKTYR 68  
 QY 105 LDAGYSKDELAETKCKVMSQLESEFVLYTRAGYGRKGEKDLTHGYIYAIPYL 164  
 DB 69 MAYSQVDELMAKRCATLTAKNNLTSAIIRPLVFIQ-DVGMGVNPPGYNTPYITAAFW- 126  
 QY 165 WAFPPAEQI-FGTTAIVPRHVRAGNTVDPTIK--NYQWDLTPAASEAKRDGARTAI 220  
 DB 127 GAVLGALEALGQIDAVVSSM-NRVAAPTITPAKAGNTLSLIVGS--EARRHGQEG 183  
 QY 221 LDDSDNCVAEGPFWNCIVKDKLASP--SRNALGCTRTKTFELADMGIEATLRDVT 278  
 DB 184 ADVHGYLSEAGENLFEVKDGLIFTPPTSSALPGITDAILAKDKMGLEVRQVLSR 243  
 QY 279 RELYDADELMAVTTAGVTPINSIDGAVNGEPGLTVAIHDRF-----WALM 327  
 DB 244 ESITLADVDFMNGTAEITPVSVDIQVIGIKRGRPVAKIQOAFPGLTGTEDKMGWL 303  
 QY 328 DEGP 332  
 DB 304 DQVNP 308

RESULT 11

Q9HNF8 PRELIMINARY; PRT; 308 AA.

AC Q9HNF8;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Branched-chain amino acid aminotransferase.  
 GN lIve2 OR VNG2122G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OC NCB1\_TaxID-64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-20504483; PubMed-11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorson V., Shogma J.,  
 RA Swartzell S., Welt D., Hall J., Dahl T.A., Welt R., Goo Y.A.,  
 RA Leitner B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
 RA Madocks D.G., Jablonski P.E., Krebs M.P., Angewine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AE005101; AAG20262.1; -.  
 DR HSP; P00510; IAG3.  
 DR Interpro; IPR001544; AminoTran\_4.  
 DR Pfam; PF01063; aminoTran\_4; 1.  
 DR ProDom; PD001961; AminoTran\_4; 1.  
 DR TIGRfams; TIGR01122; lIve\_I; 1.  
 DR Transferrase; AminoTran; Complete proteome.  
 KM Transferrase; Complete proteome.  
 SQ SEQUENCE 308 AA; 34209 MW; C86A5408B6D93A12 CRC64;

Query Match 16.9%; Score 298; DB 17; Length 308;

Best Local Similarity 27.7%; Pred. No. 5.1e-16;  
Matches 93; Conservative 56; Mismatches 127; Indels 60; Gaps 12;

QY 32 YSDYEDTSSPPAGVNAIEGEEYLPAAEAKISIDTGGHSDLYTVAHVHGN---IF 87  
DB 4 FSEMDVDT-----IMWDEGFVMEQVAVHVLTHALHYSGIFEGVRYDTDNGPAIF 55  
QY 88 RLGDHDLRLDGLASRLDAGYSKDELAETKCKVMSQLESFVNLTGTYG---GKRK 144  
DB 56 RMDHDLQFYFESAKRYDIDFETPELDATVELLQRRDLSCVIRPLATYGSLESGSP 115  
QY 145 GKKDLSKLTGHOVYIATYLAFFPAEQIFGTALVPRHVRAGRNVDPTIKNYQ---- 200  
DB 116 GD-----CPTDYALAMP--WG-----AYLGDDAL-----ENGVDVAVSTWRKHAS 154  
QY 201 -----WGDLTAASFACDGRARFALILDSNCAVEGPFNCIYKDGKLASP- 247  
DB 155 SOIPTNATKTGLVNSMLAGEBARNGFTETALVLRKSNVAGEGENTFVRDGLFTPG 214  
QY 248 -SRNALPGITRKTYVELADQMGTEATLRD---VTSRELYDADELMAYTTAGVPPINSLD 303  
DB 215 LSESLIDGITRDTYVITLAEERGY--TVHADVSIKRGELHTADELFFGSAEVTPIKQVD 272  
QY 304 GEAVNGEPPETVAIRGRFALMDEPPLIETLEY 339  
DB 273 NVEIGSGTRGPVTELEQTAFDVLV-EAGDRBEMPHY 307

## RESULT 12

Q9KW26 PRELIMINARY; PRT; 282 AA.  
AC Q9KW26;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE D-alanine aminotransferase.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
RT Staphylococcus.  
RX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SMRAIN-COI;  
RX MEDLINE=20031141; PubMed=1056865;  
RA de Lencastre H., Wu S.W., Pinho M.G., Ludovice A.M., Filipe S.,  
Gardete S., Sobral R., Gill S., Chung M., Tomasz A.;  
RT "Antibiotic resistance as a stress response: complete sequencing of a  
large number of chromosomal loci in Staphylococcus aureus strain COL  
RT that impact on the expression of resistance to methicillin";  
RL Microb. Drug Resist. 5:163-175(1999).  
DR EMBL: Y14816; CAB82475.1; -;  
DR HSSP: P19938; IDAA.  
DR InterPro: IPR001544; AminoTran\_4.  
DR Pfam: PF01063; aminoTran\_4; 1.  
DR ProDom: PD001961; AminoTran\_4; 1.  
DR TIGRFAWS: TIGR01121; D.amino.aminoT. 1.  
DR PROSITE: PS00770; AA-TRANSFER-CLASS\_4; UNKNOWN\_1.  
KW AminoTransferase; Transferase.  
SQ SEQUENCE 282 AA; 31894 MW; 6B13D0B038699A0F CRC64;

Query Match 16.5%; Score 291; DB 2; Length 282;  
Best Local Similarity 27.1%; Pred. No. 1.7e-15;  
Matches 78; Conservative 54; Mismatches 124; Indels 32; Gaps 6;

QY 49 WISEEYLPAAEAKISIDTGGHSDLYTVAHVHGNIFRLGDHDLRLDGLASRLDAG 108  
DB 5 FLNDFEYSPSAKSYNDRGVFGDGIYERYNGKLTFTYETHEYERFLSANDEIGLDN 64  
QY 109 YSKDELAETTKCVMSQLESFVNLTGKRGKRGKDELSKLTGHOVYIATYLAFF 168  
DB 65 YSVEELIELSKRLVDNQIETGAIYIQAIRGAVERNHNSFPEVEPAIVATYSYDRPD 124  
QY 169 PABQIFGTALVPRHVRAGRNVDPTIKNYQ-----GDLTAASFACDGRAR 217

DB 125 HLEN-----GVNGV--TVEDIRMLRCDIKSLNGLVLAKEVAAYKNAVE 167

QY 218 TAILLSDNCVAGPGFNVCIVADGKL-ASPSRN-ALPGITRKTYVELADQMGTEATLRD 275  
DB 168 A--IQRGETVTEGSSSNMAYAKDGIYTHPIPNNTILNCTIRIVIKKIAEDYNIPKEBT 225

QY 276 VTSRELYDADELMAYTTAGVPPINSLDGEAVNGEPPETVAIRDRF 323  
DB 226 FTYDFELKNADVEIVSTSAEVTPIYKLDGEPVADKVGFTIRQLDGEF 273

## RESULT 13

Q99TB4 PRELIMINARY; PRT; 282 AA.  
AC Q99TB4;  
DT 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE D-alanine aminotransferase.  
GN SAV1750 OR SA1571.  
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
OS Staphylococcus aureus (strain N315).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
RT Staphylococcus.  
RX NCBI\_TaxID=158878, 158879;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-S.aureus (strain Mu50), and S.aureus (strain N315);  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iino J.-Q., Ito T.,  
Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,  
Mizutani-Ul Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
Sekiizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,  
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
RT aureus";  
RL Lancet 357:1225-1240(2001).  
DR EMBL: AP003363; BAB57912.1; -;  
DR EMBL: AP003135; BAB42839.1; -;  
DR HSSP: P19938; IDAA.  
DR InterPro: IPR001544; AminoTran\_4.  
DR Pfam: PF01063; aminoTran\_4; 1.  
DR ProDom: PD001961; AminoTran\_4; 1.  
DR TIGRFAWS: TIGR01121; D.amino.aminoT. 1.  
DR PROSITE: PS00770; AA-TRANSFER-CLASS\_4; UNKNOWN\_1.  
KW Transferase; AminoTransferase; Complete proteome.  
SQ SEQUENCE 282 AA; 31908 MW; 6B0D215F38699A0F CRC64;

Query Match 16.4%; Score 290; DB 16; Length 282;  
Best Local Similarity 26.7%; Pred. No. 2e-15;  
Matches 77; Conservative 55; Mismatches 124; Indels 32; Gaps 6;

QY 49 WISEEYLPAAEAKISIDTGGHSDLYTVAHVHGNIFRLGDHDLRLDGLASRLDAG 108  
DB 5 FLNDFEYSPSAKSYNDRGVFGDGIYERYNGKLTFTYETHEYERFLSANDEIGLDN 64  
QY 109 YSKDELAETTKCVMSQLESFVNLTGKRGKRGKDELSKLTGHOVYIATYLAFF 168  
DB 65 YSVEELIELSKRLVDNQIETGAIYIQAIRGAVERNHNSFPEVEPAIVATYSYDRPD 124  
QY 169 PABQIFGTALVPRHVRAGRNVDPTIKNYQ-----GDLTAASFACDGRAR 217  
DB 125 HLEN-----GVNGV--TVEDIRMLRCDIKSLNGLVLAKEVAAYKNAVE 167  
QY 218 TAILLSDNCVAGPGFNVCIVADGKL-ASPSRN-ALPGITRKTYVELADQMGTEATLRD 275  
DB 168 A--IQRGETVTEGSSSNMAYAKDGIYTHPIPNNTILNCTIRIVIKKIAEDYNIPKEBT 225  
QY 276 VTSRELYDADELMAYTTAGVPPINSLDGEAVNGEPPETVAIRDRF 323





GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:36:48 ; Search time 15 seconds  
(without alignments)

664,958 Million cell updates/sec

Title: US-10-067-291-1

Sequence: 1 MRLSDGLGSLNLAVERGAI.....KDRFWALDEPGLIETLEY 339

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCOTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	339	4	US-09-527-522-1
2	385	21.8	288	4	US-09-173-300-29
3	322.5	18.3	310	4	US-09-173-300-31
4	315	17.9	303	2	US-08-599-171A-32
5	315	17.9	303	2	US-08-646-590B-32
6	315	17.9	303	3	US-09-069-226-32
7	315	17.9	303	4	US-09-412-184-32
8	289	16.4	289	3	US-08-972-902-6
9	283	16.0	282	3	US-08-972-902-7
10	278.5	15.8	283	3	US-08-972-902-9
11	250	14.2	283	3	US-08-972-902-8
12	250	14.2	283	4	US-09-466-257A-8
13	250	14.2	283	4	US-08-424-797A-3
14	245	13.9	283	4	US-09-466-257A-10
15	245	13.9	283	4	US-09-134-001C-5100
16	244	13.8	283	1	US-08-723-896-3
17	238	13.5	285	4	US-09-173-300-24
18	226.5	12.8	255	4	US-09-173-300-33
19	222	12.6	297	4	US-09-173-300-26
20	165.5	9.4	362	4	US-09-134-001C-5403
21	163.5	9.3	348	4	US-09-173-300-19
22	161.5	9.2	243	4	US-09-173-300-22
23	154.5	8.8	307	4	US-09-173-300-15
24	146	8.3	198	4	US-09-173-300-11
25	139.5	7.9	180	4	US-09-173-300-28
26	110.5	6.3	115	4	US-09-173-300-13
27	102.5	5.8	363	4	US-09-173-300-20

28	100	5.7	19	4	US-09-527-522-5	Sequence 5, Appl1
29	99.5	5.6	607	2	US-08-472-534-5	Sequence 5, Appl1
30	91	5.2	2860	2	US-08-826-267-2	Sequence 2, Appl1
31	89	5.0	625	1	US-08-365-981-13	Sequence 13, Appl1
32	88.5	5.0	4551	3	US-09-320-878-1	Sequence 1, Appl1
33	88.5	5.0	4613	4	US-09-105-537-31	Sequence 31, Appl1
34	88.5	5.0	11877	4	US-09-105-537-6	Sequence 6, Appl1
35	88	5.0	287	1	US-08-365-981-9	Sequence 9, Appl1
36	86.5	4.9	380	2	US-08-846-762-86	Sequence 86, Appl1
37	86.5	4.9	491	4	US-09-134-001C-4727	Sequence 4727, Ap
38	85	4.8	413	2	US-08-282-197C-49	Sequence 49, Appl1
39	82.5	4.7	303	4	US-09-134-001C-4256	Sequence 4256, Ap
40	81.5	4.6	600	4	US-09-352-159-36	Sequence 36, Appl1
41	81.5	4.6	600	4	US-09-352-159-38	Sequence 38, Appl1
42	81	4.6	700	1	US-07-720-589-2	Sequence 2, Appl1
43	81	4.6	700	2	US-08-785-190-2	Sequence 66, Appl1
44	81	4.6	700	4	US-08-235-836C-66	Sequence 2, Appl1
45	81	4.6	700	5	PCT-US92-05539-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-09-527-522-1  
; Sequence 1, Application US/09527522  
; Patent No. 6413752  
; GENERAL INFORMATION:  
; APPLICANT: Takashima, Yoshiki  
; TITLE OF INVENTION: Protein capable of catalyzing transamination  
; TITLE OF INVENTION: stereoselectively, gene encoding said protein  
; FILE REFERENCE: 058251  
; CURRENT APPLICATION NUMBER: US/09527,522  
; CURRENT FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: JP 11/075511  
; PRIOR FILING DATE: 1999-03-19  
; PRIOR APPLICATION NUMBER: JP 11/088634  
; PRIOR FILING DATE: 1999-03-30  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 339  
; TYPE: PRT  
; ORGANISM: Mycobacterium aurum  
; FEATURE:  
; OTHER INFORMATION: SC-S423  
US-09-527-522-1

Query Match 100.0%; Score 1764; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 3.9e-191;  
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MRLSDGLGSLNLAVERGAIREDTPAGSVIOYSDVELDSSFPAGVAMEEYLPAEEA	60
DB	1	MRLSDGLGSLNLAVERGAIREDTPAGSVIOYSDVELDSSFPAGVAMEEYLPAEEA	60
QY	61	KISIFPTGSHSLDTVTVAHVHGNIFRGLDHLRLDLGASRLRLDAGYSKDELAETRK	120
DB	61	KISIFPTGSHSLDTVTVAHVHGNIFRGLDHLRLDLGASRLRLDAGYSKDELAETRK	120
QY	121	CVMSQLRSEFVNLVTRGKGRKGRKDKSLKTHOYIYAIYLAFFPAEIOFGTTAIV	180
DB	121	CVMSQLRSEFVNLVTRGKGRKGRKDKSLKTHOYIYAIYLAFFPAEIOFGTTAIV	180
QY	181	PRHVRAGRTVDPPTKNTQMGDLTAASFEAKDRGARTAILDSDNCVAGGFNVCIYK	240
DB	181	PRHVRAGRTVDPPTKNTQMGDLTAASFEAKDRGARTAILDSDNCVAGGFNVCIYK	240
QY	241	DGKLAPSRNALPGITRKTVFELADQMGIEATLRDVTSHRELYDADBLMAVTTAGVTPIN	300
DB	241	DGKLAPSRNALPGITRKTVFELADQMGIEATLRDVTSHRELYDADBLMAVTTAGVTPIN	300

**QY**    301 SLDEGAVNGEGPGLTVAIDRFALMDEGPPIETIEY    339  
       |||||  
**Db**    301 SLDEGAVNGEGPGLTVAIDRFALMDEGPPIETIEY    339

RESULT 2  
DS-09-17

```

US-09-173-300-29
; Sequence 29, Application US/09173300
; Patent No. 6451581
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hiltz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT APPLICATION NUMBER: US/09/173,300
; CURRENT FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: 60/063,423
; EARLIER FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 29
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Methanococcus jannaschii
US-09-173-300-29

```

Query Match	21.88;	Score 385;	DB 4;	Length 288;
Best Local Similarity	31.98;	Pred No 4	2a-35.	

[illegible]

### RESULT 3

US-09-173-300-31  
Sequence 31, Application US/09173300  
Patent No. 6451581  
GENERAL INFORMATION:  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Hltz, William D.  
APPLICANT: Kinney, Anthony J.  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Rafalski, J. Antoni  
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES  
FILE REFERENCE: BB-116  
CURRENT APPLICATION NUMBER: US/09/173,300  
CURRENT FILING DATE: 1998-10-15  
EARLIER APPLICATION NUMBER: 60/063,423  
EARLIER FILING DATE: 1997 October 28  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Microsoft Word Version 7.0A

```

; SEQ ID NO 31
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Zea mays
US-09-173-300-31

```

Query Match	18.38;	Score 322.5;	DB 4;	Length 310;
Best Local Similarity	29.08;	Pred. No. 5.7e-28;		
Matches 82;	Conservative 54;	Mismatches 134;	Indels 13;	Gaps 5;

```

OY 47 VAMIEBEYIPAEBAEISIFPDGFHSIDLYTVVAWHMGNIFRGHIDRLDQASKLTD 106
OY 18 LVMWGBELLPRMSAKVAFSDYVQGGQAVMEGLRIIDQKFKLDEHIDRLDPSAKMAAFS 77
OY 107 AGISKUELAE-ITKCVSMQSLRESFVNLYVTGKGRKGEKDSKLTHQYIYAIPYL- 164
Db 78 NVPTROMIDALFEKTLANGFNENNAHILRLTLR-----GKRYTQSMSPAFNLGCALIV 131
OY 165 ---WAPPEAQIFGTALYVPRHVRAGRNVDPTIKNYQMGDLTASFEARDGARGATIL 221
Db 132 LAEMKRPVYDNSHG-IKLYVTATRRRNSPNSIDPKRIHNNLILNILLAKIEGULAQEDAIM 190
OY 222 LBSDNCAVBGPFCVNCIYKDGKSLASPSRN-ALDPITRKTVFELADOMGEATLDDYTSRE 280
Db 191 LDMDGVSETNATNFEWKKIVLTLPADYCLDGLITRATVMDLVYKENVFLHERISISE 250
OY 281 LYDADELMAVTAGGVPTINSLDGEANGNEBPPLVATIRDRP 323
Db 251 FHAADDEVWTTGIMGILTYVWIMIDREIGDGRIGVTRQJOKAY 293

```

## RESULT

US-08-599-171A-32  
; Sequence 32, Application US/08599171A  
: Patent No. 5814473

```

GENERAL INFORMATION:
APPLICANT: WARREN, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARLLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,171A
FILING DATE: Concurrently
CLASSIFICATION: 435
Prior APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HERRON, CHARLES J.
REGISTRATION NUMBER: 28, 019
REFERENCE/DOCKET NUMBER: 331400-38
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 AMINO ACIDS
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN

```







; SEQUENCE CHARACTERISTICS:

CLASS/DENOS: Linear  
TOPOLOGY: linear

REPORT: LINEAR





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 17:39:37 ; Search time 20 Seconds

(without alignments)  
1281.655 Million cell updates/sec

Title: US-10-067-291-1

Perfect score: 1764  
Sequence: 1 MTFALSDIGTSLNVAVEPGA1.....RDRFWALMDPEGLIETIEY 339

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 7561385 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications, AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1764	100.0	339	US-10-067-291-1	Sequence 1, Appl1
2	847	48.0	324	US-10-112-706-1	Sequence 1, Appl1
3	847	48.0	325	US-10-112-706-2	Sequence 2, Appl1
4	847	48.0	329	US-10-112-706-3	Sequence 3, Appl1
5	847	48.0	330	US-10-112-706-4	Sequence 4, Appl1
6	385	21.8	288	US-10-027-450-29	Sequence 29, Appl1
7	322.5	18.3	310	US-10-027-450-31	Sequence 31, Appl1
8	315	17.9	303	US-10-060-432-32	Sequence 32, Appl1
9	315	17.9	303	US-09-905-173-32	Sequence 32, Appl1
10	289	16.4	289	US-10-136-253-6	Sequence 7, Appl1
11	283	16.0	282	US-10-136-253-7	Sequence 7, Appl1
12	278.5	15.8	283	US-10-136-253-9	Sequence 9, Appl1
13	250	14.2	283	US-10-136-253-8	Sequence 8, Appl1
14	238	13.5	285	US-10-027-450-24	Sequence 24, Appl1
15	226.5	12.8	255	US-10-027-450-33	Sequence 33, Appl1
16	222	12.6	297	US-10-027-450-26	Sequence 26, Appl1
17	169.5	9.6	179	US-09-734-017A-42	Sequence 42, Appl1
18	163.5	9.3	348	US-10-027-450-19	Sequence 19, Appl1
19	161.5	9.2	243	US-10-027-450-22	Sequence 22, Appl1

20	154.5	8.8	307	US-10-027-450-15	Sequence 15, Appl1
21	146	8.3	198	US-10-027-450-11	Sequence 11, Appl1
22	139.5	7.9	180	US-10-027-450-28	Sequence 28, Appl1
23	137	7.8	131	US-08-781-986A-5224	Sequence 5224, Ap
24	133.5	7.6	379	US-09-738-626-5914	Sequence 5914, Ap
25	122	6.9	314	US-09-738-626-6338	Sequence 6338, Ap
26	110.3	6.3	115	US-10-027-450-13	Sequence 13, Appl1
27	108	6.1	440	US-09-734-369-178	Sequence 178, Appl
28	102.5	5.8	363	US-10-027-450-20	Sequence 12, Appl1
29	100.5	5.7	410	US-09-746-660A-12	Sequence 12, Appl1
30	100.5	5.7	456	US-09-746-660A-10	Sequence 10, Appl1
31	100	5.7	19	US-10-067-291-5	Sequence 5, Appl1
32	99.5	5.6	607	US-09-815-242-13379	Sequence 13379, A
33	99.5	5.6	607	US-09-815-242-13682	Sequence 13682, A
34	96	5.4	488	US-09-738-626-6181	Sequence 6181, Ap
35	93	5.3	782	US-09-813-408-29	Sequence 29, Appl1
36	91	5.2	721	US-10-025-187-2	Sequence 2, Appl1
37	89	5.0	554	US-09-738-626-6289	Sequence 6289, Ap
38	88.5	5.0	664	US-09-815-242-13820	Sequence 13820, A
39	88.5	5.0	4613	US-09-860-846-31	Sequence 31, Appl1
40	88.5	5.0	4613	US-09-861-289-31	Sequence 31, Appl1
41	88.5	5.0	11877	US-09-860-846-6	Sequence 6, Appl1
42	88.5	5.0	11877	US-09-861-289-6	Sequence 6, Appl1
43	85.5	4.8	996	US-09-738-626-3847	Sequence 3847, Ap
44	85	4.8	1047	US-09-866-582-37	Sequence 37, Appl1
45	84.5	4.8	620	US-09-815-242-11472	Sequence 11472, A

## ALIGNMENTS

```

RESULT 1
US-10-067-291-1
; Sequence 1, Application US/10067291
; Patent No. US20020127664A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Satoshi
; TITLE OF INVENTION: Protein capable of catalyzing transamination
; TITLE OF INVENTION: stereoselectively, gene encoding said protein
; FILE REFERENCE: 058251
; CURRENT APPLICATION NUMBER: US/10/067,291
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 09/527,522
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: JP 11/075511
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: JP 11/088634
; PRIOR FILING DATE: 1999-03-30
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Mycobacterium aurum
; FEATURE:
; OTHER INFORMATION: SC-S423
US-10-067-291-1
Query Match 100.0%; Score 1764; DB 12; Length 339;
Best Local Similarity 100.0%; Pred. No. 2.6e-159;
Matches 339; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTFALSDIGTSLNVAVEPGA1REDTPAGSVIOYSDELDTSSPPAGGVAMIEGELYRAEEA 60
Db 1 MTFALSDIGTSLNVAVEGA1REDTPAGSVIOYSDELDTSSPPAGGVAMIEGELYRAEEA 60
QY 61 KISTFDGFGHSDLTLYVAHVHGNINIRLDGHDRLDLDGASKRLDAGYSKDELAETTK 120
Db 61 KISTFDGFGHSDLTLYVAHVHGNINIRLDGHDRLDLDGASKRLDAGYSKDELAETTK 120
QY 121 CVMSQSLRESFVNLTVRGGRKGEKLSKTHQVYIYAIPIYIMAPPAEQIFGTAAIV 180

```

```

Db 121 CVSMQSLRSEFNVLTGKYGKREKDLKLTQVYIYALPYLMAFPAPQIGETAIY 180
QY 181 PRRVRAGNTVDPTIKYQMGDLTAASFEAKRCARFALLDSNCAEGPFWVCVK 240
Db 181 PRRVRAGNTVDPTIKYQMGDLTAASFEAKRCARFALLDSNCAEGPFWVCVK 240
QY 241 DGLASPSRNALPGITRKTVELADQMGIEATLRDVTRELYDADELMATVTTAGVPEIN 300
Db 241 DGLASPSRNALPGITRKTVELADQMGIEATLRDVTRELYDADELMATVTTAGVPEIN 300
QY 301 SLDEAVNGEPGPLTVAIRDRFALMDPEPLIETIY 339
Db 301 SLDEAVNGEPGPLTVAIRDRFALMDPEPLIETIY 339

```

```

RESULT 2
US-10-112-706-1
; Sequence 1, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1yuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhito
; APPLICANT: OGURA, Masahito
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
; US-10-112-706-1

```

```

Query Match 48.0%; Score 847; DB 9; Length 324;
Best Local Similarity 51.0%; Pred. No. 2,6e-72;
Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

```

```

QY 30 IQSDYELDTSSPPAGVAMIEGEYLPAEAKISIFDTGFGHSDLTYYTAHVHNGNIRL 89
Db 15 ITYSDYELDPANPLAGAAAMIEGAFVPPSEARISIFDQGLHSDVYTVFHVWNGNAERL 74
QY 90 GDHLRLDGLSKRLDAGYSKDELAETRKCVSMQSLRSEFNVLTGKYGKREKDL 149
Db 75 DDHIERLFESNAESMRILPLOTDEVEKIALELVAKTELRAFAVSITRGYSSTGERDI 134
QY 150 SKLTHQVYIYALPYLMAFPAPQIGETAIYPRHVRAGNTVDPTIKYQMGDLTAASF 209
Db 135 TKHRPOVYIMAVPQWTVFDRIRDGVAHMAVSRTPRSSIDPOVKNFQMGDLIRAVQ 194
QY 210 EAKRCARFALLDSNCAEGPFWVCVKDKGLASPSRNALPGITRKTVELADQMG 269
Db 195 ETHDRGEAFRLDLDGGLAEGSGFNVVYIKDGVVSPGAAALPGITRKTVELASLGH 254
QY 270 EATLRVTSRELYDADELMATVTTAGVTPINSLDGEAVNGEPGPLTVAIRDRFALMD 329
Db 255 EATLRDITLAEILDADDEVLGCTTAGVMPFVSDGNPISDGVPGBITQSIIRRYWELNVE 314
QY 330 PGPLEIETIY 339
Db 315 SSSLTPVOY 324

```

```

RESULT 3
US-10-112-706-2
; Sequence 2, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1yuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhito
; APPLICANT: OGURA, Masahito
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: 09/403,493
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: JP 9-121732
; PRIOR FILING DATE: 1997-04-23
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Arthrobacter sp.
; US-10-112-706-2

```

```

Query Match 48.0%; Score 847; DB 9; Length 325;
Best Local Similarity 51.0%; Pred. No. 2,6e-72;
Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

```

```

QY 30 IQSDYELDTSSPPAGVAMIEGEYLPAEAKISIFDTGFGHSDLTYYTAHVHNGNIRL 89
Db 16 ITYSDYELDPANPLAGAAAMIEGAFVPPSEARISIFDQGLHSDVYTVFHVWNGNAERL 75
QY 90 GDHLRLDGLSKRLDAGYSKDELAETRKCVSMQSLRSEFNVLTGKYGKREKDL 149
Db 76 DDHIERLFESNAESMRILPLOTDEVEKIALELVAKTELRAFAVSITRGYSSTGERDI 135
QY 150 SKLTHQVYIYALPYLMAFPAPQIGETAIYPRHVRAGNTVDPTIKYQMGDLTAASF 209
Db 136 TKHRPOVYIMAVPQWTVFDRIRDGVAHMAVSRTPRSSIDPOVKNFQMGDLIRAVQ 195
QY 210 EAKRCARFALLDSNCAEGPFWVCVKDKGLASPSRNALPGITRKTVELADQMG 269
Db 196 ETHDRGEAFRLDLDGGLAEGSGFNVVYIKDGVVSPGAAALPGITRKTVELASLGH 255
QY 270 EATLRVTSRELYDADELMATVTTAGVTPINSLDGEAVNGEPGPLTVAIRDRFALMD 329
Db 256 EATLRDITLAEILDADDEVLGCTTAGVMPFVSDGNPISDGVPGBITQSIIRRYWELNVE 315
QY 330 PGPLEIETIY 339
Db 316 SSSLTPVOY 325

```

```

RESULT 4
US-10-112-706-3
; Sequence 3, Application US/10112706
; Publication No. US20020192786A1
; GENERAL INFORMATION:
; APPLICANT: YAMADA, Yukio
; APPLICANT: IWASAKI, Akira
; APPLICANT: KIZAKI, No. US20020192786A1yuki
; APPLICANT: MATSUMOTO, Keiji
; APPLICANT: IKENAKA, Yasuhito
; APPLICANT: OGURA, Masahito
; APPLICANT: HASEGAWA, Junzo
; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
; FILE REFERENCE: 1422-398P
; CURRENT APPLICATION NUMBER: US/10/112,706
; CURRENT FILING DATE: 2002-04-02

```

PRIOR APPLICATION NUMBER: 09/403,493  
 PRIOR FILING DATE: 1999-10-22  
 PRIOR APPLICATION NUMBER: JP 9-121732  
 PRIOR FILING DATE: 1997-04-23  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: Patentia version 3.0  
 SEQ ID NO 3  
 LENGTH: 329  
 TYPE: PRT  
 ORGANISM: *Arthrobacter* sp.  
 US-10-112-706-3

Query Match 48.0%; Score 847; DB 9; Length 329;  
 Best Local Similarity 51.0%; Pred. No. 2,6e-72;  
 Matches 156; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

QY 30 IQYSDYELDTSSPPAGVAMIEGTYLPAEAKISIFDTGFGHSDLTYYVAHWGNIFRL 89  
 DB 20 ITYSDELDPANPLAGGAMIEGAFVPPSEARISIFDGYLHSDVTVFVHWGNIFRL 79  
 QY 90 GDHLDRLDGAASKLRDAGYSKDELAETTKCVMSOLRESFVNLTVRGYGRKGEKDL 149  
 DB 80 DDHIERLFSNAESMKRIPLTQDEYKELELVAKTELEAFVSITRGYSTPGEKDI 139  
 QY 150 SKLHQVYIYAIPLYLWAPPAEIOFGTTAIVPRHVRAGRTVDTIKNYQMGDLTAASF 209  
 DB 140 TKHROPVYMAVYQWIVFDRIRGVHMAVQSVRRTPRSSIDPQVKNFQMGDLIRAVQ 199  
 QY 210 EAKDGAFTAILDSDNCVABGPFVNCIVDGLASRNALPGITRTVFEILDQNGI 269  
 DB 200 EYHDSGFAPPLLDGGLAEGSGFNVAIKDGVRSGRALPGITRTVFEILDQNGI 259  
 QY 270 EATLDVTSRELYDADELMVATAGVTPINSIDGAVNGEPPGLTVAIRFVALMDE 329  
 DB 260 EATLDITLAEILDADDEVLCCTTAGGVWPFVSDGNIPISDVGPPITQSIIRRYWELNVE 319  
 QY 330 PGPLIETIEY 339  
 DB 320 SSSLTPVQY 329

RESULT 5  
 US-10-112-706-4  
 Sequence 4, Application US/10112706  
 Publication No. US20020192786A1  
 GENERAL INFORMATION:  
 APPLICANT: YAMADA, Yukio  
 APPLICANT: IMASAKI, Akiro  
 APPLICANT: KIZAKI, No. US20020192786A1YUKI  
 APPLICANT: MATSUMOTO, Keiji  
 APPLICANT: IKENAKA, Yasuhiko  
 APPLICANT: OGURA, Masahiro  
 APPLICANT: HASEGAWA, Junzo  
 TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds  
 FILE REFERENCE: 1442-398P  
 CURRENT FILING DATE: 2002-04-02  
 PRIOR APPLICATION NUMBER: US/10/112,706  
 PRIOR FILING DATE: 1999-10-22  
 PRIOR APPLICATION NUMBER: JP 9-121732  
 PRIOR FILING DATE: 1997-04-23  
 NUMBER OF SEQ ID NOS: 26  
 SOFTWARE: Patentia version 3.0  
 SEQ ID NO 4  
 LENGTH: 330  
 TYPE: PRT  
 ORGANISM: *Arthrobacter* sp.  
 US-10-112-706-4

Query Match 48.0%; Score 847; DB 9; Length 330;  
 Best Local Similarity 51.0%; Pred. No. 2,7e-72;  
 Matches 158; Conservative 56; Mismatches 96; Indels 0; Gaps 0;

QY 30 IQYSDYELDTSSPPAGVAMIEGTYLPAEAKISIFDTGFGHSDLTYYVAHWGNIFRL 89  
 DB 21 ITYSDELDPANPLAGGAMIEGAFVPPSEARISIFDGYLHSDVTVFVHWGNIFRL 80  
 QY 90 GDHLDRLDGAASKLRDAGYSKDELAETTKCVMSOLRESFVNLTVRGYGRKGEKDL 149  
 DB 81 DDHIERLFSNAESMKRIPLTQDEYKELELVAKTELEAFVSITRGYSTPGEKDI 140  
 QY 150 SKLHQVYIYAIPLYLWAPPAEIOFGTTAIVPRHVRAGRTVDTIKNYQMGDLTAASF 209  
 DB 141 TKHROPVYMAVYQWIVFDRIRGVHMAVQSVRRTPRSSIDPQVKNFQMGDLIRAVQ 200  
 QY 210 EAKDGAFTAILDSDNCVABGPFVNCIVKOKLASPSRNALPGITRTVFEILDQNGI 269  
 DB 201 EYHDSGFAPPLLDGGLAEGSGFNVAIKDGVRSGRALPGITRTVFEILDQNGI 260  
 QY 270 EATLDVTSRELYDADELMVATAGVTPINSIDGAVNGEPPGLTVAIRFVALMDE 329  
 DB 261 EATLDITLAEILDADDEVLCCTTAGGVWPFVSDGNIPISDVGPPITQSIIRRYWELNVE 320  
 QY 330 PGPLIETIEY 339  
 DB 321 SSSLTPVQY 330

RESULT 6  
 US-10-027-450-29  
 Sequence 29, Application US/10027450  
 Patent No. US20020102715A1  
 GENERAL INFORMATION:  
 APPLICANT: Falco, Saverio Carl  
 APPLICANT: Hiltz, William D.  
 APPLICANT: Kinney, Anthony J.  
 APPLICANT: Cahoon, Rebecca E.  
 APPLICANT: Rafalski, J. Antoni  
 TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES  
 FILE REFERENCE: BB-1126  
 CURRENT FILING DATE: 2001-12-20  
 PRIOR APPLICATION NUMBER: 60/063,423  
 PRIOR FILING DATE: 1997 October 28  
 NUMBER OF SEQ ID NOS: 54  
 SOFTWARE: Microsoft Word Version 7.0A  
 SEQ ID NO 29  
 LENGTH: 288  
 TYPE: PRT  
 ORGANISM: *Methanococcus jannaschii*  
 US-10-027-450-29

Query Match 21.8%; Score 385; DB 12; Length 288;  
 Best Local Similarity 31.9%; Pred. No. 1,5e-28;  
 Matches 90; Conservative 58; Mismatches 116; Indels 18; Gaps 5;

QY 49 WIEGEYLPABERAKISIFDTGFGHSDLTYYVAHWGNIFRLGDHLDRLDGAASKLRDAG 108  
 DB 4 YLNGKFYDEKDAKVSVDHGLLYDGVFEGIRADVFMKEHIDRLYSASLSICDIP 63  
 QY 109 YSKDELAETTKCVMSOLRESFVNLTVRGY-----KRGEDLSKLNHOVYIYAI 163  
 DB 64 LTKEMIDVYLETIRVNNLRDARTLRVTVTRGVGLDPRKCGP-----TICLALP- 116  
 QY 164 LMAFPAPFOIGTTAIVPRHVRAGRTVDTIKNYQMGDLTAASFPAKRGARTAILD 223  
 DB 117 ---MPRLGEGIRALIV-SYRRLPYDLNPAVSLNYSVLAKIQANVAGVDEAFILD 172  
 QY 224 SDNCVABGPFVNCIVKOKLASPS--RNALPGITRTVFEILDQNGIETLRVTSREL 281  
 DB 173 DKGFVEEGTDNIRIVKNGVAKTPPVYQSILKGTIRVVYIKIAEEGIEVVEEPLTLHDL 232  
 QY 282 YDADELMVATAGVTPINSIDGAVNGEPPGLTVAIRDRF 323  
 DB 233 YTADELITGTAAELVVFIEDGKVINNKOVGEITKLEKRF 274

RESULT 7  
US-10-027-450-31  
Sequence 31, Application US/10027450  
Patent No. US20020102715A1  
GENERAL INFORMATION:  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Hiteo, William D.  
APPLICANT: Kinney, Anthony J.  
APPLICANT: Cahoon, Rebecca E.  
APPLICANT: Rafalski, J. Antoni  
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES  
FILE REFERENCE: BB-1126  
CURRENT APPLICATION NUMBER: US/10/027,450  
CURRENT FILING DATE: 2001-12-20  
PRIOR APPLICATION NUMBER: 60/063,423  
PRIOR FILING DATE: 1997 October 28  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: Microsoft Word Version 7.0A  
SEQ ID NO 31  
LENGTH: 310  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-027-450-31

Query Match 18.3%; Score 322.5; DB 12; Length 310;  
Best Local Similarity 29.0%; Pred. No. 1.5e-22;  
Matches 82; Conservative 54; Mismatches 134; Indels 13; Gaps 5;  
QY 47 VAMIEGEVLPAAEAKISIFDGFSGHSDLYTVAHVHNGN--IFRLGDHLDRLL 106  
DB 18 LVWGVDELLPNSAKVSVFSDVVOGDAVWEGRLTYDKVFKDEHDLRLEDSAKAMAFS 77  
QY 107 AGYSKDELAE-ITKCVSMQSLRESFVNLVTRGSGRKEKDKSLTHOVIYIAPYL- 164  
DB 78 NVPRDMIKDAIFKTLANGFMENNAHRLTLTR-----GKVTSGSPAFNLIGCALLY 131  
QY 165 ---WAPPARDIFGTALVRRHVARAGRNVDPIKAYONGDILTAASFKAQKARTAIL 221  
DB 132 LAEKPPVYDISHG-ILVATATRRNSPNSIDPKIHNNLNNILAKIEGNLAQAEDAIM 190  
QY 222 LDSNCAVEGFGFVNCIVKDKKLASPSRN-ALPGITRTKTYFELADOMGIEATLRDVTSS 280  
DB 191 LDKGPFSENNATNIFVVKGVILTPHADCLGIFRTATYMDLVKENVLHERRISLSE 250  
QY 281 LYDADELAAYTTAGCVTPINSLSGEAVNGEPPPLVAIDRF 323  
DB 251 FHADEVWTTGTMGELTPVVMIDGREIGDKIGPVTROIQKAY 293

RESULT 8  
US-10-060-432-32  
Sequence 32, Application US/10060432  
Publication No. US20030040092A1  
GENERAL INFORMATION:  
APPLICANT: WARREN, Patrick V.  
APPLICANT: SWANSON, Ronald V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/060,432  
FILING DATE: 29-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/481,733  
FILING DATE: 11-JAN-2000  
APPLICATION NUMBER: US/08/599,171  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: HERRON, CHARLES J.  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 331400-38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 AMINO ACIDS  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
SEQUENCE DESCRIPTION: SEQ ID NO: 32:  
US-10-060-432-32

Query Match 17.9%; Score 315; DB 9; Length 303;  
Best Local Similarity 28.0%; Pred. No. 7.3e-22;  
Matches 85; Conservative 62; Mismatches 113; Indels 44; Gaps 11;  
QY 42 PRFGVAMIEGEVLPAAEAKISIFDGFSGHSDLYTVAHVHNGN--IFRLGDHLDRLL 97  
DB 3 PYAKYI-WLDGRILKEDAKIHVTLHALHGTSTIFGNGYNGNDLTVRLREHIDMAY 61  
QY 98 DGASKRLDAGYSKDELAETITKCVSMQSLRE-----SEV-NLTVTRGSGRKEKEDL 149  
DB 62 RSKATIGIPIPTREVRVAVLETTKANFRDVIYIRPAFAVASQVTT-----LDI 112  
QY 150 SKLTHOVIYIAPYLWAPPARDIFGTALVRRHVR-----RAGRNVDPPIKNGY--W 201  
DB 113 RNLVSLAVTVP-----FG-KYLSPNGIKATIVSRHRVHNTLMPEMAKIGIT 160  
QY 202 GDLTASFPAKRGARTAILDSDNCVABEGFVNCIVKDKLSP--SRNALPGITRT 259  
DB 161 VNSVTLAVKARSRGDEALMDVNGYVEGSGENLFIYRGGLFTPPVHESILLEGITRD 220  
QY 260 VVELADOMGIEATLRDVSRELVDDELMAVTTAGCVTPINSLSGEAVNGEPPPLTAI 319  
DB 221 VAKLSGDVGLRVEEKPIITREYVTTADEVFLVGTAEITPVEVDGRTIGTKPGPIITKI 280  
QY 320 RDRE 323  
DB 281 RELY 284

RESULT 9  
US-09-905-173-32  
Sequence 32, Application US/09050173  
Patent No. US20020132295A1  
GENERAL INFORMATION:  
APPLICANT: DIVERSA CORPORATION  
APPLICANT: SHORT, Jay M.  
APPLICANT: WARREN, Patrick V.  
APPLICANT: SWANSON, Ronald V.  
APPLICANT: MATHUR, Elie J.  
TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND  
FILE REFERENCE: DIVER1240-7  
CURRENT APPLICATION NUMBER: US/09/905,173  
CURRENT FILING DATE: 2001-07-12  
PRIOR APPLICATION NUMBER: US 09/412,184  
PRIOR FILING DATE: 1999-10-04  
PRIOR APPLICATION NUMBER: US 09/389,537  
PRIOR FILING DATE: 1999-09-02  
PRIOR APPLICATION NUMBER: US 08/646,590



MALCHIES	03,	COMBETVALVE	21,	MISMAIACHES	120,	LIUGED	30,	SUPP
----------	-----	-------------	-----	-------------	------	--------	-----	------

US-10-136-253-9

```

; Sequence 9, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Fred R.
; APPLICANT: PORTNOY, Daniel A.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
; FILE REFERENCE: 053893-5011-02
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US/10/136,253
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: PCT/US98/24357
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 08/972,902
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-136-253-9

```

```

Query Match          15.8%; Score 278.5; DB 12; Length 283;
Best Local Similarity 27.6%; Pred. No. 1.9e-18;
Matches 83; Conservative 50; Mismatches 131; Indels 37; Gaps 9;

```

```

OY 45 GGVAMIEGELPAEFAKISIFDTGFGHSDLTYYVAHWGNIFRLGDLRLDGASKLR 104
DB 2 GYTLN-NQIYVDEKVIDKEDRGYQFGDGYEVYKYNNGEMFYNEHIDLYIAAEKIR 60
OY 105 LDAGYSKDELAETTKCVSMQSLRESFVNLVTRGCGRKGEKLSKLTHOVYIYALPYL 164
DB 61 ITIPYTKKFKHQLHELVKNEKMLNGHIYFQYTRGTSR-----AHQ----- 102
OY 165 WAFPP---AEQFETTAIVPRHVRAGNTVDPTIKNQW---GDUTA-----ASREA 211
DB 103 --FPPNTKPVITGYTKNPPLENLEKGVKATPEVEDIRMLRCIDKISLNLGAVLAKQEA 160
OY 212 KDRGARTAILDSDNCVAGPGFNVCYKDKGL-ASPSRN-ALPGITRTVFEELADOMGI 269
DB 161 HEKGYEAI-LHRNTVTEGSSSNVFGIKDKILYHPANNILIGITRDVYIACANEINM 219
OY 270 EATLDSVSRLEYDADELAATVTAAGVTPINSLSGEAVNGEPPLVAIRDREWALMDE 329
DB 220 PVKELEPFTTHALKMDELFTVSTSEITPVLEIDGKLIRDKGVEMWTKLQKOFETKIPK 279
OY 330 P 330
DB 280 P 280

```

```

RESULT 13
US-10-136-253-8
; Sequence 8, Application US/10136253
; Patent No. US20020136737A1
; GENERAL INFORMATION:
; APPLICANT: FRANKEL, Fred R.
; APPLICANT: PORTNOY, Daniel A.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACIDS COMPRISING LISTERIA DAL AND DAT GENES
; FILE REFERENCE: 053893-5011-02
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US/10/136,253
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: PCT/US98/24357
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: 08/972,902
; PRIOR FILING DATE: 1997-11-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 283

```

```

; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-136-253-8

```

```

Query Match          14.2%; Score 250; DB 12; Length 283;
Best Local Similarity 26.9%; Pred. No. 9.7e-16;
Matches 79; Conservative 49; Mismatches 110; Indels 56; Gaps 9;

```

```

OY 58 EEAISISIF--DTGFGHSDLTYYVAHWGNIFRLGDLRLDGASKLRDLADAGSYDELA 115
DB 12 EEGSITSPEDRGYQFGDGYEVYKYNNGEMFYNEHIDLYIAAEKIRLVYITVDVH 71
OY 116 EITKCVSMQSLRESFVNLVTRGCGRKGEKLSKLTHOVYIYALPYLWAFPPAQIRG 175
DB 72 KLHLDLEKNNLNNGHIYFQYTRGTSR-----HIFEDA-----S 107
OY 176 TTAIVPRHVRAGNTVDPTIKNQW-----DLTA-----ASREAND 213
DB 108 VPAVLTVGNV---TGERSIENFEGVKATLVEDYRWMLRCIDKISLNLGAVLAKQEA 162
OY 214 RGARTAILDSDNCVAGPGFNVCYKDKGL-ASPSRN-ALPGITRTVFEELADOMGIER 271
DB 163 KGCYEALIHGSD-ITPCCSANYGIRDKGLYHPANNILIGITRDVYIACANEINLV 221
OY 272 TLRDVTSRELYDADELAATVTAAGVTPINSLSGEAVNGEPPLVAIRDREWA 325
DB 222 IEEPNTKGDLLTMDIELIVSSVSSEVTPYIDVDQGLGAGVPGSEWTKRLKAKAPA 275

```

```

RESULT 14
US-10-027-450-24
; Sequence 24, Application US/10027450
; Patent No. US20020102715A1
; GENERAL INFORMATION:
; APPLICANT: Falco, Saverio Carl
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, J. Antoni
; TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
; FILE REFERENCE: BB-1126
; CURRENT FILING DATE: US/10/027,450
; CURRENT APPLICATION NUMBER: 2001-12-20
; PRIOR FILING DATE: 60/063,423
; PRIOR FILING DATE: 1997 October 28
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Microsoft Word Version 7.0A
; SEQ ID NO 24
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Glycine max
US-10-027-450-24

```

```

Query Match          13.5%; Score 238; DB 12; Length 285;
Best Local Similarity 28.2%; Pred. No. 1.3e-14;
Matches 81; Conservative 48; Mismatches 116; Indels 42; Gaps 10;

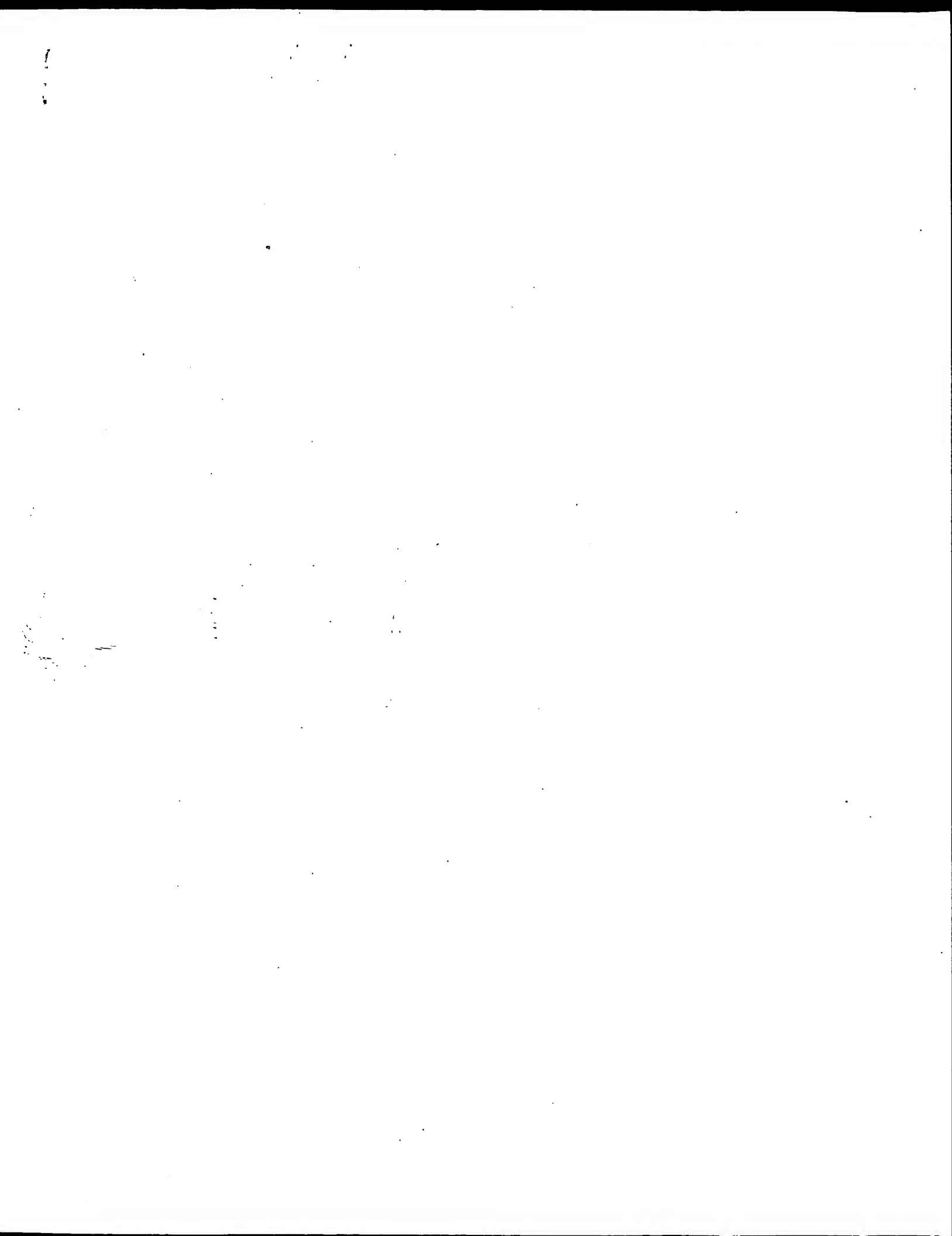
```

```

OY 51 EGEYLPAAEAKI-----SIFDTGFGHSDLTYYVAHWGNIFRLGDLRLDGASKLRDL 106
DB 2 EAAVITPMDDHMAVRGHGVDT-----NALMDGLVLELDQHLDFLTSASMSKID 50
OY 107 AGYSKDELAETTKCVSMQSLRESFVNLVTRGCGRKGEKLSKLTHOVYIYALPYLM 165
DB 51 PPDRODSIRRLILQYTSASKRCRGLRYVLSAG-----PDDPFLDSGCCRRSLYAIYOD 106
OY 166 APPPAQIRG---TTAIVPRHVRAGNTVD--PTIKNQWGDILTAASFEAKRGART 219
DB 107 LSPSSNPFNGVYKVVYSSIPKHPKFAITKSVYLPV-----LSKVEAEAGAFVG 157
OY 220 IILDSDNCVAGPGFNVCYKDKGLASPS-RNALPGITRTVFEELADOMGIEAL---- 273
DB 158 IWLDSGEFVAGBNMVAFTVTKDELIMPHDKILSGCTAKRYVLTAASLAREGLKIGIR 217

```





GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 20:06:42 ; Search time 2799 Seconds

(without alignments)  
10605.524 Million cell updates/sec

Title: US-10-067-291-2

Perfect score: 1020  
Sequence: 1 atgactgccttcagacct.....tcgaacgatacgaatcga 1020Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenBankl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vl:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vl:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rtd:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	100.0	1020	BD000505	BD000505 Protein,
2	247	24.2	329709	AP002997	AP002997 Mesorhizo
C 3	191.6	18.8	307551	AP003015	AP003015 Mesorhizo
C 4	153.2	15.0	149050	ML0672114	ML0672114 Mesorhizo
C 5	129.4	12.7	10364	AE005957	AE005957 Caulobact
C 6	126.6	12.4	10685	AE010451	AE010451 Methanopy
C 7	109.8	10.8	190050	AL646080	AL646080 Ralstonia
C 8	95.8	9.4	12730	AE005101	AE005101 Halobacte
C 9	95	9.3	348077	AP003000	AP003000 Mesorhizo
C 10	87.2	8.5	212050	AL646060	AL646060 Ralstonia
C 11	85.8	8.4	17576	AE004913	AE004913 Pseudomon
C 12	84	8.2	10993	AE011153	AE011153 Methanosa
C 13	83.2	8.2	7991	PA063816	PA063816 Pseudomonas
C 14	82.6	8.1	10196	AE004996	AE004996 Halobacte
C 15	72.4	7.1	10029	AE013330	AE013330 Methanosa
C 16	72.4	7.1	23533	AE001039	AE001039 Archaeogl
C 17	70	6.9	2748	SAUSICA	M94370 Stigmatella
C 18	68.8	6.7	12026	AE005109	AE005109 Halobacte
C 19	68.6	6.7	31226	SC67	AL353870 Streptomy
C 20	67.8	6.6	14991	AE005125	AE005125 Halobacte
C 21	67.2	6.6	347660	AP002994	AP002994 Mesorhizo
C 22	64.6	6.3	11202	AE005123	AE005123 Halobacte
C 23	64.2	6.3	12510	AE005848	AE005848 Caulobact
C 24	64.2	6.3	77457	AF210249	AF210249 Streptomy
C 25	63.2	6.2	30000	AX250261	AX250261 Sequence
C 26	62.8	6.2	1173	AF246405	AF246405 Acidithio
C 27	62.8	6.2	11170	AE005047	AE005047 Halobacte
C 28	62.8	6.2	40544	SC5	AE005366 Streptomy
C 29	62.2	6.1	10112	AE012915	AE012915 Chlorobiu
C 30	62.2	6.1	23404	SC10A9	AL583943 Streptomy
C 31	62	6.1	20107	AE008882	AE008882 Salmonell
C 32	62	6.1	96086	STSTMD1	AF233324 Salmonell
C 33	61.2	6.0	45313	SCD95A	AL374332 Streptomy
C 34	60.8	6.0	2475	AF039028	AF039028 Streptomy
C 35	60.8	6.0	15901	SC3A3	AL109849 Streptomy
C 36	60.8	6.0	32039	SC7A1	AL034447 Streptomy
C 37	60.8	6.0	69301	STU82965	U82965 Streptomyce
C 38	60.4	5.9	10652	AB006206	AB006206 Streptomy
C 39	60.4	5.9	34593	SC2K8	AB089164 Streptomy
C 40	60.2	5.9	6835	FS053363	U53363 Frankia sp.
C 41	60	5.9	23030	SC9E12	AL391751 Streptomy
C 42	59.6	5.8	3522	AF421216	AF421216 Streptomy
C 43	59.4	5.8	31624	SCD63	AL161755 Streptomy
C 44	59.2	5.8	54101	AB070956	AB070956 Streptomy
C 45	58.4	5.7	9169	AX427931	AX427931 Sequence

## ALIGNMENTS

RESULT 1  
BD000505  
LOCUS BD000505 1020 bp DNA  
DEFINITION Protein, its gene and use thereof.  
ACCESSION BD000505  
VERSION BD000505.1 GI:18623618  
KEYWORDS JP 2000342276-A/1.  
SOURCE unidentifed.  
ORGANISM unidentifed.  
REFERENCE 1 (bases 1 to 1020)  
AUTHORS Takashima,Y., Wieser,M., and Koda,K.  
TITLE Protein, its gene and use thereof  
JOURNAL Patent: JP 2000342276-A 1 12-DEC-2000;  
SUMITOMO CHEMICAL CO LTD

Pred. No. is the number of results predicted by chance to have a

COMMENT OS Mycobacterium aurum SC-S423  
 PN JP 2000342276-A/1  
 PD 12-DEC-2000  
 PR 17-MAR-2000 JP 2000075752  
 PI YOSHIMI TAKASHIMA, MARKOS WIESER, KEN KODA  
 PC C12N15/09, C07K14/35, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC  
 C12P13/00,  
 CC C12P21/02//C07M7:00, C12N15/00, C12N5/00  
 C12P13/00,  
 CC C12P21/02//C07M7:00, C12N15/00, C12N5/00

FEATURES  
 source Location/Qualifiers  
 1.1020  
 /organism="unidentified"  
 /db\_xref="taxon:32644"

BASE COUNT 198 a 350 c 313 g 159 t  
 ORIGIN

Query Match 100.0%; Score 1020; DB 6; Length 1020;  
 Best Local Similarity 100.0%; Pred. No. 7e-131;  
 Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGCTCTTTAGACCTTCGACCTCCAACTGCTGCGCCGCGCCGATC 60  
 DB 1 ATGACGCTCTTTAGACCTTCGACCTCCAACTGCTGCGCCGCGCCGATC 60  
 QY 61 CGCGAGACACCCCGCGGCTCGGTATCCAGTACAGCAGTACGAACTGACCTTC 120  
 DB 61 CGCGAGACACCCCGCGGCTCGGTATCCAGTACAGCAGTACGAACTGACCTTC 120  
 QY 121 AGCCGCTTCCCGCGGCGCTCGGTATCCAGTACAGCAGTACGAACTGACCTTC 180  
 DB 121 AGCCGCTTCCCGCGGCGCTCGGTATCCAGTACAGCAGTACGAACTGACCTTC 180  
 QY 181 AAGATCTCATCTTTCGACACCGGATTCGATTCGATTCGATTCGATTCGATTCGATTC 240  
 DB 181 AAGATCTCATCTTTCGACACCGGATTCGATTCGATTCGATTCGATTCGATTCGATTC 240  
 QY 241 GATGACACGCGCAACATCTTCGCGCTCGCGGACACCTGACCGGATTCGATTCGATTC 300  
 DB 241 GATGACACGCGCAACATCTTCGCGCTCGCGGACACCTGACCGGATTCGATTCGATTC 300  
 QY 301 TCCAGTGGCGCTCGACGCGCGGATTCAGACGAGAGAGAACTGCGCGGATTCAGACG 360  
 DB 301 TCCAGTGGCGCTCGACGCGCGGATTCAGACGAGAGAGAACTGCGCGGATTCAGACG 360  
 QY 361 TGGGTGTCGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 420  
 DB 361 TGGGTGTCGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 420  
 QY 421 GGAAGCGCAAGGCGGAG 480  
 DB 421 GGAAGCGCAAGGCGGAG 480  
 QY 481 ATCCGCTACCTGCGGCTTCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 DB 481 ATCCGCTACCTGCGGCTTCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 QY 541 CCGCGCATGTCG 600  
 DB 541 CCGCGCATGTCG 600  
 QY 601 TGGGTGTCGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 660  
 DB 601 TGGGTGTCGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 660  
 QY 661 CTGCTGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 720  
 DB 661 CTGCTGATTCGACGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 720  
 QY 721 GACGGCAAGCTGCGCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 DB 721 GACGGCAAGCTGCGCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780

DB 721 GACGGCAAGCTGCGCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 QY 781 TTGGAACCTGGCCGACCAAGATGGGATATGAGGACCACTTCGCGGAGTCAACGCGGTGA 840  
 DB 781 TTGGAACCTGGCCGACCAAGATGGGATATGAGGACCACTTCGCGGAGTCAACGCGGTGA 840  
 QY 841 CTCTAGACGCGGAGAGAGTGTGATGGCGGACCACTTCGCGGAGTCAACGCGGTGA 900  
 DB 841 CTCTAGACGCGGAGAGAGTGTGATGGCGGACCACTTCGCGGAGTCAACGCGGTGA 900  
 QY 901 TCGCTGATGAGGAGGCGCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 901 TCGCTGATGAGGAGGCGCTTCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 QY 961 GACCGGTTTGGGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
 DB 961 GACCGGTTTGGGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

RESULT 2  
 LOCUS Ap002997 329709 bp DNA linear BCT 15-MAY-2001  
 DEFINITION Mesorhizobium loti DNA, complete genome, section 4/21.  
 ACCESSION AP002997 BA000012  
 VERSION AP002997.2 GI:14022051  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mesorhizobium loti (strain:MAFF303099) DNA.  
 Mesorhizobium loti  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Phyllobacteriaceae; Mesorhizobium.  
 1 (sites)  
 Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,  
 Watanabe, A., Idegawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,  
 Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,  
 Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimp, S., Sugimoto, M.,  
 Takeuchi, C., Yamada, M. and Tabata, S.  
 Complete genome structure of the nitrogen-fixing symbiotic  
 bacterium Mesorhizobium loti  
 DNA Res. 7 (6), 331-338 (2000).  
 2 (bases 1 to 329709)  
 Kaneko, T.  
 Direct Submission  
 Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research, Yana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:kaneko@kazusa.or.jp,  
 URL:http://www.kazusa.or.jp/rhizobase/  
 Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
 On May 11, 2001 this sequence version replaced gi:11994965.

COMMENT  
 FEATURES  
 Location/Qualifiers  
 1.329709  
 /organism="Mesorhizobium loti"  
 /strain="MAFF303099"  
 /db\_xref="taxon:381"  
 complement(79..402)  
 /gene="ml1124"  
 complement(79..402)  
 /gene="ml1124"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48662.1"  
 /db\_xref="GI:14022052"  
 /translation="MSRKGIISIPFNSPHYVADCLMVPPIISDLSIFPAFTMR  
 LKGPRLRPQVLAITGILAVVAMLLNLNLMPIYDDDFVIGVTFISWMAALFDA  
 ILIA"  
 411..2480  
 /gene="mlr1245"  
 411..2480  
 /gene="mlr1245"  
 /codon\_start=1  
 /transl\_table=11

/product="probable electron transfer protein"  
 /protein\_id="BAB48663.1"  
 /db\_xref="GI:14022053"  
 /translation="MNSPANITDPLVLPFGSKRGKRPVGVLDAAOLGVYESVC  
 GGRATCGGQIEVGEVGFANFKHIVSSNHSIPKGEKERYERVGLPERRRLSGSAOI  
 LGLDIYDIPQDVINAOTIRKDAORVIAKPTAIKMCVLEEDPMHPLCIDLRKI  
 ALMKMGKRNLEFDYTLPOVGLIKSNKWTATAIHKDADSDIARVIALPGLKNEA  
 YGLACDGGSTTAMHLVSLSGRVASSGTSNPQIRGEDLSRVSYVMNPDEKREK  
 TVAVREASISLVDKVCAGNVQRNDILDSVGVNIMHFLGIDPTELGAAPALAV  
 SAVERIKASDGLKNGCARLYMLPCIGHVGAADAAVTLSEGPPODEMMLIYDGT  
 NAEVLGNRRARYAASSPTGPAFEGAEISGGORAPAEIERYRIDPTLEKRYVIGS  
 ELMSDEPGLSVQATGYTGICSGSIIITVAEMYLACIISDGVDDSLARSRYTA  
 NERTSYLKEBSEPKITITQDVRAIQAKALAGTILMEKQNTEDVDRHKGAP  
 GSEIDPKTAMVGLIPDCDLKVSANAGANAPALLNNGYREIETYSQJEKE  
 TALEPKFQEHFYAMALPNKYDPPFKLSAAVKLPKRTVSEBGIAAGDAPRRSRSEH  
 AARSRE"  
 gene  
 2529. 2852  
 /gene="mlr1246"  
 2529. 2852  
 /gene="mlr1246"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48664.1"  
 /db\_xref="GI:14022054"  
 /translation="MPPRIDPETETLFSRNLSSTWQAHLPRIERREPIILGRSSRE  
 NKMSEFENIGRGVTRKQNHAKNAIRLNSLPAHVQADIGWPASPPSDPALASL  
 LGTAR"  
 gene  
 3087. 4040  
 /gene="mlr1247"  
 3087. 4040  
 /gene="mlr1247"  
 /codon\_start=1  
 /transl\_table=11  
 /product="esterase"  
 /protein\_id="BAB48665.1"  
 /db\_xref="GI:14022055"  
 /translation="MPSLKSHEVSVLVAISRRKAPSSPENLRMTAAARKTEHHPPA  
 SLQRLHIOQTRVDFPVYEIAPKAGEORILLYLHGAAYVEITPYHMLIADMAADR  
 GYGIYVPIPIAEPHDFHAFMGWGVYROMIDEADIVFMGDSAGNNAVLTVM  
 AAEGILPLPARHLISPGIDMSLNPKLFEARNDEPMGLAGEALIEMYAGIDRS  
 WHISPIGVSVLPTTLTSGSHDLTPDNLEFQKADADVEYVVEHEEOMFAMPL  
 IDMPARRARQIIVAFLEGNRPVQDQKTRFQASAAAE"  
 gene  
 complement(4062. 4961)  
 /gene="ml11248"  
 complement(4062. 4961)  
 /gene="ml11248"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional regulator"  
 /protein\_id="BAB48666.1"  
 /db\_xref="GI:14022056"  
 /translation="MNAFLNHPILPLDLVLTAEVAIAETGSETTANAVAPRTSVA  
 NOIKLEDLIGRSVADARASVLTDEMLGIRLLINREVSVFIIPEIYGV  
 RIGSPDDYGERVLPVLRKPAOSHSPIAVDTIIOSSSLRRRMDRALDITLLNSYK  
 TSNLSAEVLTREPIYMAKAGCAGLRRPLPSISMEBCANRAGALELAGEGRYRY  
 ATMSAHTAGORAIMADLAVAPLPSFLGNDVVELCPKQMDIGITINLAVVAPDAS  
 AVKAVADHIRATFEVFERETGE"  
 gene  
 5113. 5379  
 /gene="mr1249"  
 5113. 5379  
 /gene="mr1249"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48667.1"  
 /db\_xref="GI:14022057"  
 /translation="MGLKHAACKTSQIKRRNVPAPAKENEMSTLSUGRIATESAA  
 RAKYOTERAIRSLPIELQKIDICWPASDPTKGTTRNGVSWAGAK"  
 gene  
 6136. 7137  
 /gene="mlr1251"  
 6136. 7137  
 /gene="mlr1251"  
 /gene="mlr1251"

/codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional regulator"  
 /protein\_id="BAB48668.1"  
 /db\_xref="GI:14022058"  
 /translation="MNAIKPIKRSFVPELVDPFTMAFAATDPLRSANRMGYEY  
 RMRLASIDGKPVASNCVCAVNTSLEEEKKAGDPKPMALVCSGVNERYONKSA  
 FAWLREYNNGVAAGGCTGAHLLAAGLLSNRCALHWNELPGSEAPKAVFADL  
 FEVDQNTVTCAGGAALDMMKLIGDDEDSLVNRCVCEVLTDRVSPDROKPLPA  
 RLGVONKVTYIETMEGNSEPLSLIEIDHVDLSRQILELFFEMGRSPARYILE  
 IRLDRARHLIIOSSMPVEYVAVACGFSAHSFKCYREIYARSPQDERVDRQLAA"  
 gene  
 complement(7153. 7956)  
 /gene="ml11253"  
 complement(7153. 7956)  
 /gene="ml11253"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP-binding protein of ABC transporter"  
 /protein\_id="BAB48669.1"  
 /db\_xref="GI:14022059"  
 /translation="MEPVSKLLIEGVSRTFAGVGGI.PYKALMPIDLAANDFTTI  
 LGPSCGCKSTLIRIVAGLEAPSEGRVYLDKRAVTRPQPDGVAFQSYTLFPMVSEN  
 IAFGLRGNAPKREDDIVASVYDVLGKFEENHMPQISGGMOORAILARLANDE  
 ILLDERGALDNOTRGMLKEILLIGWERRKTYLPTHDIEEAIMASRYWTARP  
 GSISKDAVIDPHEPRHTYLTSPSEFSLKARLIEDIRVEAMRTAQAP"  
 gene  
 complement(7941. 8720)  
 /gene="ml11254"  
 complement(7941. 8720)  
 /gene="ml11254"  
 /codon\_start=1  
 /transl\_table=11  
 /product="permease protein of ABC transporter"  
 /protein\_id="BAB48670.1"  
 /db\_xref="GI:14022060"  
 /translation="MKMRPLPVSPGLRTGLISFVYLFIAFMATILGSHVNRIFLA  
 DPLSMLRDMRLVEDREFMILITTRVGVGVLSVAVPGLIMGAKPVEAFLE  
 PVSFAPVYIPASAFPIILITMAGIGELKILYFNGSVPTIILIVAKVSTRDVE  
 AATYLGSTNGINARVITPMAPEIATLVLGMATYIVAVELLGSSSGIGITMIIN  
 SOSILATGQIILFGIITVIGLISDFEAKFNNWLPWLSIA"  
 gene  
 complement(8802. 9752)  
 /gene="ml11256"  
 complement(8802. 9752)  
 /gene="ml11256"  
 /note="hypothetical protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48671.1"  
 /db\_xref="GI:14022061"  
 /translation="MNRIVAFATIFPAGLASAPVMAATSVTIGISGTFEAPILAK  
 QAGCFERHGLDYTLKRYPAQSRPLAISGLQCAITTEVFLWMAVNSGVTTOITFD  
 KSYGADIVVRNDIKTYADLKGKVAASAGTSPTFLAVLNRKNGSTADTVNLE  
 PDAQAQAFIAGONDAAVTEPFISAVDKADQGHITLTDYPMVLDVTCTEFLKA  
 NPDAKALADSYFALDLIKDKPOKSYEIGADVOKSAKFEEDSAVYIKWADRADNOQ  
 FTFEPDQSTAGALLQGLIKREAPDVATLADTSVAN"  
 gene  
 complement(9663. 11000)  
 /gene="ml11257"  
 complement(9663. 11000)  
 /gene="ml11257"  
 /note="unknown protein"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB48672.1"  
 /db\_xref="GI:14022062"  
 /translation="MSSISARPIKRSKSPVATSNFMTRTSSRAADRSGPRDLQKP  
 GFEPGCLFKOVVAYPRGMSALSRAHLISGCAALAMLMMPGSRADBPVVEVLV  
 AVDVLSMSADELEIQRHGYAALITHBNVLAADAGHGLIATVYVMAATGTORVYV  
 PMWTANRADERVIAOLSAOPSPASARTISGLLEGGSLPESGEYGERKVIIDISG  
 DGPNOGAPVNTFEDGYRGCIYNGIPLMTFRGISAYAVVNDLDRYSQVYGGERA  
 FMIPVNTQPEPRIRKKLVLELAGPASPOMAAEABADHPVVLITQDPAADCOIGEMK  
 WRNNMMLDSR"  
 gene  
 complement(10906. 13467)  
 /gene="ml11258"  
 complement(10906. 13467)







DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
Db 184630	GTGTAGCGGCTCTCCGGACACAAAGTCGGCCATTGTGGAAATGATCTGTACGGCAGG---TG		184574				
OY 422	GAAGCGCAGAGGCGCAGAAAGACCTGTCCAGCTCAACCATCAGGTGTACATCTACGCCA		481				
Db 184573	GCTGCACCCACACTTAAAGCCGGGACCCGCGCAATCGAAGAACGGTTTCAATCGATTCCGGG		184514				
OY 482	TCCCGTACCTTGTGGGCTCTTCCCGCCCGCAGACATATCTTGGACACCAACCGGATCTGTC		541				
Db 184513	TGCCCTTTGGGGTCGG---TTGCCAACAAAGAGACACTCGAGGCGCGCTCGATGTTCGCCA		184457				
OY 542	CGCGCCATGTCCGCCCGCCGCGCCGACACACCCTGTGCACCCGCATCAAGAACTACCACT		601				
Db 184456	TCAGCAACACTGTCCCGGATCCCGCCCAAGTCATGATCCGACATCCGAAACTATCTACT		184397				
OY 602	GGGGATCTTCACCGCAGCAATGTTTCGAAGCCAAAGACCGGTGTCGGCGGACCGCGATCC		661				
Db 184396	GGCTCGATCTGGTGAAGGGCTCTTTCGATGCTACGACTATGTGTCCGAACCTGATTTGA		184337				
OY 652	TGCTCGACTCGGACACTGTGCGTGGCCGGAAGTCCGGGCTTCAACGTGTGATCGTCAAG		721				
Db 184336	TGCTGACATCAACACAAACATTTGCCAAAGGCGCTGGCTTAACTTTACCGCTCAAG		184277				
OY 722	ACGGCAGCTGGGCTCTCCCGTCCCGGAACGCTTCCGGGATCAACCCGTAAAGAGGTGT		781				
Db 184276	ACGGCGCTTGAAGACGCCGGTGTATGTGTCTTTCGGCGGACTTACAGGACACCGTAT		184217				
OY 782	TGCACTGGCCACACAGATGGGCAATCGAAGCCACCTCTCGCAGCTCAACACCGCTGAC		841				
Db 184216	TGTATCTGTGGACGAACTGGGCTTATCTGTACGGCGCGGCGCACTTGATGCAACGAAAC		184157				
OY 842	TCTACGAGCGCGACAGATTGATGGCGTCCACCACCGGGGGGGTCAACCCGATCAACT		901				
Db 184156	TGAAGGGGCGCCATAGGTTTTCATCACGTTCACAGACGGGGATCATGTCCGGTGTCCA		184097				
OY 902	CGCTGATGGGACGCGCTGGGACACGCGACCCCGGTACACTGACGGTGGCCATCCGG		961				
Db 184096	AGATTGACGAACCTTTGTAGGCGACAGCGCAAGTTGGCGCTCTAACACGCCAGCTGGCAG		184037				
OY 962	ACCGGTTCTGGGCGCTGATGACGAGCGCG 991						
Db 184036	ACCTGTATTGGGAGAACGATGCTGTATCCG 184007						
RESULT 4							
LOCUS	ML0672114/c	149050 bp	DNA	linear	BC1 14-MAY-2002		
DEFINITION	Mesophilobium loti R7A symbiosis island; segment 3/4.						
ACCESSION	AL672114 AL672111						
VERSION	AL672114.1	GI:20804067					
KEYWORDS							
SOURCE	Mesophilobium loti.						
ORGANISM	Mesophilobium loti Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Phyllobacteriaceae; Mesophilobium.						
REFERENCE	1 (bases 1 to 149050)						
AUTHORS	Sullivan,J.T., Trzebiatowski,J.R., Cruickshank,R.W., Gouzy,J., Brown,S.D., Elliott,R.M., Fleetwood,D.J., McCallum,N.G., Rosbach,U., Stuart,G.S., Weaver,J.E., Webb,J.R., de Bruijn,F.J. and Ronson,C.W.						
TITLE	Comparative sequence analysis of the symbiosis island of Mesophilobium loti strain R7A						
JOURNAL	J. Bacteriol. 184 (11), 3086-3095 (2002)						
MEDLINE	21999272						
REFERENCE	12003951						
PUBMED	2 (bases 1 to 149050)						
AUTHORS	Ronson,C.W.						
TITLE	Direct Submission						
JOURNAL	Submitted (07-FEB-2002)						
COMMENT	Department of Microbiology, University of Otago, P.O. Box 56, Dunedin, New Zealand; MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824 USA; Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, Unite						

Mixte de Recherche (UMR) 215 Centre National de la Recherche Scientifique (CNRS), Institut National de la Recherche Agronomique, Chemin de Borde Rouge, BP27, F-31326 Castanet Tolosan Cedex, France, c.liverson@toulouse.inra.fr/msl.  
http://sequence.toulouse.inra.fr/msl.

location/Qualifiers  
    1. .149050  
        /organism="Mesorhizobium lotii"  
        /strain="R7A"  
        /db\_xref="taxon:381"  
        454. .1516  
        /note="Predicted by Homology"  
        /insertion\_seq="msl437"  
        520. .786  
        /gene="msl239"  
        520. .786  
        /gene="msl239"  
        /function="elements of external origin; transposon-related functions"  
        /note="Product confidence : putative  
              Gene name confidence : hypothetical"  
        /codon\_start=1  
        /evidence-not\_experimental  
        /transl\_table=11  
        /product="PUTATIVE TRANSPOSASE FOR INSERTION ELEMENT PROTEIN"  
        /protein\_id="CAD31271.1"  
        /db\_xref="GI:20804068"  
        /translation="MKRRFTBEQIIAVLRHEHGAQAGRLARKHAISEATLYWYKAK  
                    YGGMDVDAKRKLAEEDNAKKRLDLNDLMAALKELLSKK"  
        780. .1709  
        /gene="msl240"  
        780. .1709  
        /gene="msl240"  
        /function="elements of external origin; transposon-related functions"  
        /note="Product confidence : putative  
              Gene name confidence : hypothetical"  
        /codon\_start=1  
        /evidence-not\_experimental  
        /transl\_table=11  
        /product="PUTATIVE TRANSPOSASE FOR INSERTION SEQUENCE"  
        /protein\_id="CAD31272.1"  
        /db\_xref="GI:20804069"  
        /translation="WGPPARRREAVAHQAVMGLSEERRACISIVGADRKMYRNRCRP  
                    ETELRGRLRELARNERRLGYRLFILRRGEASGVNRYIRLVAREBOLTVKRRARR  
                    AVETRAPILVEARNARMSLDVFVDQSGRFRVNMIIVDVTRECIAAIPDSISGR  
                    RVAREMLDELISLGKPMIYSIDGTFTSNAIIAMSODHYEMHYIAPGMONGYIE  
                    SFNGRMDELINLEIFGLDHARCATIEWEDYNTANPHSSLSIGYQTPAAAREVLTAAG  
                    SNALDVGFASPAPAOPAPYGVTETVALIAG"  
        complement(1825. .2691)  
        /gene="msl241"  
        complement(1825. .2691)  
        /gene="msl241"  
        /function="cell processes; transport of small molecules"  
        /note="Product confidence : putative  
              Gene name confidence : putative"  
        /codon\_start=1  
        /evidence-not\_experimental  
        /transl\_table=11  
        /product="PUTATIVE AMINO-ACID ABC TRANSPORTER ATP-BINDING PROTEIN"  
        /protein\_id="CAD31273.1"  
        /db\_xref="GI:20804070"  
        /translation="MTATDDIHILGLNGDSSMATGCNVRIPIVWSOALSQOKOVPSA  
                    DKITIEATHSRWYGSFPAFLDVDLTVKRGGRIVICPSSGSGSTLIJCFNRLENBOGA  
                    EIVNGDILHNKRNATYEVRKNVGMEQHNPMTVLANNCAGPMWIGVEAOAK  
                    QTALKUYERVIRPEAOANKYEVKSGGOORVAALARSCLMPEPAVLFGEPESALDEPV  
                    SEVLETMTSLARDGMTVCVTHMGFAVAADRIFIENDKRIIVEANPNDFPTNHHE  
                    RTLFSLQSILSH"

complement(2688. .3569)  
        /gene="msl242"  
        complement(2688. .3569)

CDS



Db 8936 AACACCGCAGCTGGAAGACACAGCCTTTGGGCTGCTCCAGCATATAACGGGGAGTCCG 8897  
 779 TCTTCAGTGGCCGACACAGATGGGATCGAAGACACCTCGAGAGTACACAGCCGCG 838  
 Db 8896 TTTTGGACCTGTGTGGGAGTGGGACTTCCGGTACCGCTCGACACTCTGCTCTGG 8837  
 839 AACTCTAGACAGCCGACGATGTTGATGGCGGTACACCGCGCGGGGCTCACACCGATCA 898  
 Db 8836 AACTGGCGAAGCCGATGAGTGTTCATCAGCTCCACCGCGAGGGGGATCATGGCCGCTGA 8777  
 899 ACTCCCTGATGGCGAGCCGCTGGGCAACGGCGAGCCGGTCCACTGACGGTGGCCATCC 958  
 Db 8776 CGCGGCTGATGCTTCAATCCAGCGAGCGGTAAGTTGGCTGACCGCGAGCTGA 8717  
 959 GGGACCGGCTCTGG 972  
 Db 8716 TGGACCTCTACTCG 8703

RESULT 5  
 AE005957/c 10364 bp DNA linear BCT 12-JUN-2002  
 LOCUS Caulobacter crescentus CB15 section 283 of 359 of the complete  
 DEFINITION genome.  
 ACCESSION AE005957 AE005673  
 VERSION AE005957.1 GI:13424548  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Caulobacter crescentus CB15.  
 Caulobacter crescentus CB15.  
 Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;  
 Caulobacter.

REFERENCE  
 1 (bases 1 to 10364)  
 Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,  
 Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,  
 Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,  
 Ely,B., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,  
 Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouli,H.,  
 Shetty,J., Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,  
 Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and  
 Fraser,C.M.  
 Complete genome sequence of Caulobacter crescentus  
 Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)

TITLE  
 JOURNAL MEDLINE  
 21173698  
 PUBMED  
 11259647  
 2 (bases 1 to 10364)  
 Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,  
 Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I.,  
 Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,  
 Laub,M.T., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,  
 Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khouli,H., Shetty,J.,  
 Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathevan,J.,  
 Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and  
 Fraser,C.M.  
 Direct Submission  
 Submitted (31-JAN-2001) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 Location/Qualifiers  
 1. 10364  
 /organism="Caulobacter crescentus CB15"  
 /strain="CB15"  
 /db\_xref="taxon:190650"  
 69. 1196  
 /gene="CC2925"  
 69. 1196  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypotheical protein"  
 /protein\_id="AAK24887.1"  
 /db\_xref="GI:13424549"  
 /translation="MMLTLAALAAVAAPCPASDESPACVRRIDALPMTDLITVGTGTH

gene  
 CDS  
 69. 1196  
 /note="identified by Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypotheical protein"  
 /protein\_id="AAK24887.1"  
 /db\_xref="GI:13424551"  
 /translation="MGGRRSLSGROLPNAGRAHPAPDQGLAPKAPRLTAKMSV  
 HKWSLVSATFLMLCTIGLPVFSHELEHLMEMAPVAPADGPKLLDQVLTALA  
 RKPGVPAPMSFDEDRPVNVNYSVOPGKAGSVSPIDQSGDPAPVAPVACHPMEFL  
 LQHTDMEFLGALMGLGAMGLLILVALSGVLYAPPMRLPGTVAPSKARATRL  
 DYHNLGATVAAVAVLVGATGVNALATPIYGYMNTAKTELTAAVDSPPVAGRSGL  
 OAAVAKAOALPDQDLOVAPVAPGSDSYSDHYAVFEGHKKTELTHTTPTALDQSG  
 LAALPPTVYVTVTLSPHFGDVGAGLITLWALDIATVILASGLYMLAKRY  
 AR"  
 complement(2482. 4647)  
 /gene="CC2928"  
 complement(2482. 4647)  
 /note="CC2928"  
 /note="identified by match to protein family HMM"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TonB-dependent receptor"  
 /protein\_id="AAK24890.1"  
 /db\_xref="GI:13424552"  
 /translation="MSSRLILSALISGALASPAAROADTTRNDVYVITGQVRL  
 PPAYAGDVAAGARGLGALDMDTPPRAVTSYEBALIRNOQASVADVLDNDPTVRY  
 TKGFEDELTVYKFEVYSDDMTYNGLYVLPQVFAELIERVDVFRGNAPFLNLA  
 AFGGSGVGAFLNMPKRAPLPTLRTAGMGGRBOYGSADIAIRFGONDAVYGARVL  
 GIRGETVAVDEKRLDLYGLGLDRGGRARFSADLQODRIGARPTVPLCAIPK  
 APSAPKNAQNPWTHDEKRLDLYGLDRGGRARFSADLQODRIGARPTVPLCAIPK  
 NGTSAVREFDVRKOSLITSDAGIRAKTGAAGVHALVAVSAYOYNLSKNAYSNRA  
 GPASLTVPAVAPNPNEPTGSGSDNPVTERKNSVAVADTSLFLENRLTAVY  
 RKQDITNSIDNTGARNGATDGDVTPALAAVFKPSDKISLVANVAVLPGKILAA  
 OVNGVAVANVEILISPERGEOTEIGAKYDACTFGGSLSEFTTQPAVFEASRRYTA  
 GGEQDNQCELETVFGOPAPGRLGLGATGADQALNRLASLTIGATGIEVFEQANIN  
 LEMDVATLEGLTLEGRAVHTGSOQANTNTSLAARFIDGARYAREAGSKATVLLA  
 RVENLDRNOMVAVAGVYGANYTLTGAPRLRLISIDP"  
 complement(4868. 5758)  
 /gene="CC2929"  
 complement(4868. 5758)  
 /gene="CC2929"  
 /note="identified by match to PFM protein family HMM  
 PF01063"  
 /codon\_start=1  
 /transl\_table=11  
 /product="branched-chain amino acid aminotransferase"  
 /protein\_id="AAK24891.1"  
 /db\_xref="GI:13424553"  
 /translation="MSLVPPDDRDGMIMDQGFVPMREAVVHVLTHGLHASSFEEB  
 RMVGEIRKULEHTEERLKFSEILDEFTPTVAIDACATAKNGLKCCYVPIAM  
 RGEKMGVSAQOSKIHALVAVWMPSTYFDPATKAKGIRLTWAKYORDDPTAPILAA

AGLYMCTISKAHEKEDYADAMMLDYGVAEATGANVFVKDGLTPKPDCLDG  
 TIRRYIDAKAKGIEVERHIOKEELATFTFCFIVGTAAEVTPVSEVGEFFTPAKL  
 SIDMDTYAATVGERMAFA"  
 gene  
 5839. .6345  
 /gene="CC2930"  
 CDS  
 5839. .6345  
 /note="identified by match to protein family HMM"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator, Marr family"  
 /protein\_id="AAK24892.1"  
 /db\_xref="GI:13424554"  
 /translation="MIAPIPGSDPRLIREELDGLLEILAEASIMAAVDAVLE  
 TETLIGRSHMRAATLRRRPGIVODSKLSKQASRTLSDEKAGLVERVSGD  
 LDRRRPATLIEGVAFEORTARLALLARYRTGGDVGATRIILAAAGSHQV  
 GPERMP"  
 gene  
 6342. .7043  
 /gene="CC2931"  
 CDS  
 6342. .7043  
 /note="identified by match to protein family HMM"  
 /codon\_start=1  
 /transl\_table=11  
 /product="petr protein, putative"  
 /protein\_id="AAK24893.1"  
 /db\_xref="GI:13424555"  
 /translation="MTDPRRRLHYDDDRRLRLKEFLSRAGRYTAASAAA  
 KRLDALDPLDMVMPGEDMATYKLRKAGSGAGTPIMLTAROTDRIEGL  
 SSGVDYLKPEPEOELLRLIEALIRRSARVGPRAVLGCTEPADGELTCEBA  
 VRTAEVTLRLARSLHPVDRLADTDATGRAVDVOTRLRRKEPDKNPR  
 YLQTVRGVGRILAPD"  
 gene  
 7206. .8543  
 /gene="CC2932"  
 CDS  
 7206. .8543  
 /note="identified by match to protein family HMM"  
 /codon\_start=1  
 /transl\_table=11  
 /product="sensor histidine kinase"  
 /protein\_id="AAK24894.1"  
 /db\_xref="GI:13424556"  
 /translation="MPLTRLDIPDLKRLPTLFRSLIITLPAVIAQIAYTAF  
 DAMOSYTKLSGLADIAVAOSYEDDSPAARVAKLRAGALSLSTIAQKGL  
 PSHRPSLPAIDRLSDKLEDRDLPFMDTIRYAYIDIRVOVGGVQIYALDR  
 AYATGHIIFLAVVAITLTAVALIFIRNOVAIRLEAADAFCRGEDEPKPGA  
 RYVROALAIRKIRIORHIEORTALLASVSHDLRPLRLLEMAAEPCEMEAM  
 KGLIAEMHKIDETLAPRSGEATQVYDLSLVSVYADARGAATETETGGL  
 TRLPLTFERRALNLDINGVAHADRVATVSPKOTGVYAVDDGPGIPEDYEAF  
 KPFSRLDESRNOKVGLGLAARDMARGLGDLVLSRLALGLRLALRPG"  
 gene  
 complement (8671. .9522)  
 CDS  
 complement (8671. .9522)  
 /gene="CC2933"  
 complement (8671. .9522)  
 /note="identified by match to PFM protein family HMM"  
 /codon\_start=1  
 /transl\_table=11  
 /product="prephenate dehydratase"  
 /protein\_id="AAK24895.1"  
 /db\_xref="GI:13424557"  
 /translation="WSTLKRIAROGEGANSHERCTYPDYEARVCKTPEAPAK  
 SGVAQNLNIPLENSIAGRVAHVHLPLASGLTIGERFKIRFOJMANCKVLEDIY  
 VSSMPIALISQNSIKRLGVETEAAGTAAALAKLAPRTHAAVAPALAAAGID  
 ILIARDIDECRNITRFLMTADKAPAPDFTHRCTVSVFVRVRLPALYKALGGAT  
 NGVMTKLESMEGNTATFEVYAEVDGREDRLALALDELKFFSRFRILVYPAD  
 PFRDCA"  
 gene  
 complement (9519. .10256)  
 CDS  
 complement (9519. .10256)  
 /gene="CC2934"  
 /note="identified by match to TIGR protein family HMM"  
 TIGR00466"

Query Match 12.7% Score 129.4; DB 1; Length 10364;  
 Best Local Similarity 49.9%; Pred. No. 4.9e-09;  
 Matches 384; Conservative 0; Mismatches 376; Indels 9; Gaps 2;  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-deoxy-manno-octulosonate cytidyllyltransferase"  
 QY 132 CGGCGCGTCGCTGATCGAGGCGGAATACCTGCCGCGAAGAGCATCTCAT 191  
 DB 5732 CGACGGTTGATCTGCTTGAGCGGCAATTTGTGCCCTGGCGCGAAGAGTGCATGT 5673  
 QY 192 GTTCGACACCGGATTCGCTGATTCGAGCTACACCTGCGGCAATGATGCGACAG 251  
 DB 5672 TTGACCCATGCGCCGCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5613  
 QY 252 CAACATCTTCGCGCGGCGGACCATCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 311  
 DB 5612 CGAATCTTCAACCTGACCGGACCATCGGACCGGCTGCTGCTGCTGCTGCTGCTGCTG 5553  
 QY 312 CCGACGCGCGGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 371  
 DB 5552 TTTTGATGATCCCTACAGGCTGCGGAGATGACGAGGCTGCAAGGCGAGCGGCCAA 5493  
 QY 372 GTTCGACGCTGCGGAAATCTGCTGATCTGACCTGACCGCGGCGATACGGAACGCA 431  
 DB 5492 GAACGCTGGAAGCACTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5433  
 QY 432 GGGCGAAGAGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCT 491  
 DB 5432 TGTTCGCGCGCAGCACTGCTGCACTGCTGCACTGCTGCACTGCTGCACTGCTG 5379  
 QY 492 GTGGGCTTCG 551  
 DB 5378 CAGCTACTTCGATCCG 5319  
 QY 552 CCG 611  
 DB 5318 GCTTCG 5259  
 QY 612 CACCGACG 671  
 DB 5258 CACCATCTCCACG 5199  
 QY 672 GGAACACTGCG 731  
 DB 5198 TCCGCGCTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5139  
 QY 732 GGCCTCCG 788  
 DB 5138 GCAACG 5079  
 QY 789 GGCACGACGAGTGGGATGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848  
 DB 5078 GGCACGAGGCGGAGGAGTGGGATGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCG 5019  
 QY 849 GCGCGACGATGATGCG 897  
 DB 5018 CTTACACGAGTGGTATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4970  
 RESULT 6  
 AE010451/c 10685 bp DNA linear BCT 01-ANG-2002  
 LOCUS Methanopyrus kandleri AV19 section 150 of the complete  
 DEFINITION genome.  
 ACCESSION AE010451 AE009439  
 VERSION AE010451.1 GI:19888323  
 KEYWORDS  
 SOURCE Methanopyrus kandleri AV19.  
 ORGANISM Methanopyrus kandleri AV19.  
 Archaea; Euryarchaeota; Methanopyri; Methanopyrales;  
 Methanopyraceae; Methanopyrus.

REFERENCE	1 (bases 1 to 10685)
ATTORNS	Shesher, A.I., Mezheva, K.V., Makarova, K.S., Polushin, N.N., Shcherbina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Nale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V. and Kozaykin, S.A.
TITLE	The Complete Genome of the Hyperthermophile Methanopyrus kandleri AV19 and Monophyly of Archaeal Methanogens
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 10685)
ATTORNS	Shesher, A.I., Mezheva, K.V., Makarova, K.S., Polushin, N.N., Shcherbina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Nale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V. and Kozaykin, S.A.
TITLE	Submitted (04-FEB-2002) Fidelity Systems, Inc., Galthersburg, MD
JOURNAL	20879
FEATURES	Location/Qualifiers
source	1. 10685
	/organism="Methanopyrus kandleri AV19"
	/strain="AV19"
gene	/db_xref="taxon:190192"
	complement(123..1466)
	/gene="fifh"
	/note="MK1612"
CDS	complement(123..1466)
	/gene="ffh"
	/codon_start=1
	/transl_table=11
	/product="Signal recognition particle GTPase"
	/protein_id="AA02825.1"
	/db_xref="GI:19888324"
	/translation="MIFGRADPLAETTKIKGASIIDPEFVKEVVDVORALLDADV KIVLESKRIERALEEPPACVPRKRDILKIVBELVLSGKTEBIDILSDV VYMLGILGKGTITTAARKLARIQKRGVSVGLVADPRAPAGDQLRQLAEEVPH VEDVDVAMAVKAGVLEALDECDVIVDTAGSDRLSEDLIDELREMERIPHEVLY LDATVQKAGDAEAFHEAVOLITGVITIKDPTAKGSGALSAVANATGAPITFVGGER VDDLEFNPRESFVARLIGIDIDELIRTEEMLEEEKAEDVLSGSEPLIKLYOLEA LSKMPVKKLOYPVGMGGGNVRKISQITEERLKKYKVVINDSMTEKELNPELINS RIRRIAGSGTSEVITELNHYRMKNVVIDIOSGRIPRIGSELGRVIRNVLNG"
gene	complement(1463..2512)
	/note="f1st"
	/note="MK1613"
CDS	complement(1463..2512)
	/gene="f1st"
	/codon_start=1
	/transl_table=11
	/product="Signal recognition particle GTPase"
	/protein_id="AA02826.1"
	/db_xref="GI:19888325"
	/translation="MPSLFGKMSVLSKYKEVAKHVEERAREAKYKEVTME DLKPTLKERLKYKREVITTEADIDIDLELELISNDVAVEASIREBELKEU VGRVYKSEIIPKYVEBFRALLSVLEPKREVLMETVEKARODGRPALIMEVGNG SGTITAKVAKLKDQSVVIAAADTFRAALIEOLEEHAERLGVTLIGERDDPT AVAFNAVQHAHAKGRDVLDVDTAGRAYTDVIMELKKMKVLEPDLVIVGDALAN DAIEQAKTFHEYVGIDCAILIKVDADKAGAVISISKVTGAPILYLGVDYDILKAF SPENFEVERIGGES"
gene	complement(2542..3015)
	/gene="g1m5"
	/note="MK1614"
CDS	complement(2542..3015)
	/gene="g1m5"
	/codon_start=1
	/transl_table=11
	/product="Predicted prefoldin, molecular chaperone implicated in de novo protein folding"
	/protein_id="AA02827.1"
	/db_xref="GI:19888326"
	/translation="MAEKNQEOIQEQLRIAFINRLQCOMAINAQIDILISSTISE LNRVEETIKGVKELEGEDEVLPYGAQSFVACVDTERTVINGAGVAFRIPIDEL ESIDROELEKRAAEAOQKLOELAOELQKORAKDELALQOLGSAORIAQSSGGG"
gene	complement(3015..3257)
	/gene="f120A"
	/note="MK1615"
	complement(3015..3257)
	/gene="f120A"
	/codon_start=1
	/transl_table=11
	/product="Ribosomal protein L20A (L18A)"
	/protein_id="AA02828.1"
	/db_xref="GI:19888327"
	/translation="MSEVKVEVGRGFRMGDEPRPTROVPATSEBEALEKYSDG SEHVSMEIIEIREDIPSKVEDPILRLGLVEE"
gene	complement(3275..3940)
	/gene="f1f6"
	/note="MK1616"
CDS	complement(3275..3940)
	/gene="f1f6"
	/codon_start=1
	/transl_table=11
	/product="Translation initiation factor 6 (EIF6)"
	/protein_id="AA02829.1"
	/db_xref="GI:19888328"
	/translation="MIVKASVHGDPNIGAMIAASEYAVAVAPKPPDIVERKALD VEYRTVAGSNLVGALLAVNSGALFPRAREHEIRVVEELGVEDVLPKRNAGN LVLDHGALVHPDLDHALEVEVIGSVRGVGLGVTVGSAGVANSKGAIVHPG ATEPEMRNVEVLGVYVEVETVNRGSPYVGIVNSKGAIVGEDTTPGLARLEAL YLI"
gene	complement(3943..4203)
	/gene="f131A"
	/note="MK1617"
CDS	complement(3943..4203)
	/gene="f131A"
	/codon_start=1
	/transl_table=11
	/product="Ribosomal protein L31E"
	/protein_id="AA02830.1"
	/db_xref="GI:19888329"
	/translation="MAEVDERVYVPLDAAKAPLKKRAPYKALRPFIERHMAE EVKIDNVNKKIEMRGKIKPPSIRVRAKYADGIVEVRLAE"
gene	complement(4219..4374)
	/gene="f139"
	/note="MK1618"
CDS	complement(4219..4374)
	/gene="f139"
	/codon_start=1
	/transl_table=11
	/product="Ribosomal protein L39E"
	/protein_id="AA02831.1"
	/db_xref="GI:19888330"
	/translation="MARVPLGKKLMAKAIKQNRVPPVYAKTGRIYDNPKKRHW RRSKIKP"
gene	complement(4388..4741)
	/gene="MK1619"
CDS	complement(4388..4741)
	/gene="MK1619"
	/codon_start=1
	/transl_table=11
	/product="DNA-binding protein"
	/protein_id="AA02832.1"
	/db_xref="GI:19888331"
	/translation="MTDPELEIRIRRKIMELQRLKEESQEKVEBERKALEAOR AMRLRITPEAREBLARVRLARPOLAQAVENTLLDLAQTGLAKEXIDQLKRIILKOV SDATREVERIRRRK"
gene	complement(4771..5220)
	/gene="f151A"
	/note="MK1620"
CDS	complement(4771..5220)
	/gene="f151A"
	/codon_start=1
	/transl_table=11
	/product="Ribosomal protein S19E (S16A)"
	/protein_id="AA02833.1"
	/db_xref="GI:19888332"
	/translation="MYDAYVYVGSLEVERLAELKDPFEIRKPEMAKYVKTGRKERP PDDPMWVYRASITLRVYMDGPGVGSILRTYGGRODRCARPEREFGSAITIRKIL QDLERAGLVEKTEBGRVYTPGERSLVDSFAHIAKEKGYTTRFTSP"



```

gene      complement(5266. .5583)
           /gene="MK1621"
           complement(5266. .5583)
           /gene="MK1621"
           /transl_table=1
           /codon_start=1
           /product="Predicted RNA-binding protein containing KH
           domain, possibly ribosomal protein"
           /protein_id="AA02835.1"
           /db_xref="GI:1988833"
           /translation="MTQVKRLSGKERRALLARAVLDPVYRIKKGLITSGVIOEDVRQ
           LEEERGLIKVFERNRIIRYDRKELAEELAKVAELIDVNGRAVILFRPREGWRREHG
           LSR"
gene      complement(5570. .5956)
           /gene="rpr2"
           /note="MK1622"
           complement(5570. .5956)
           /gene="rpr2"
           /codon_start=1
           /transl_table=1
           /product="RNAse P subunit RPR2"
           /protein_id="AA02835.1"
           /db_xref="GI:1988833"
           /translation="MCKVESPRSGVLLIRIALERARLLRLARTYIEEDPDARVY
           ELARRIMAKARVLLPKHLKRSFCKRCNTPLIPGTARVRLRNMRPHVSTCLECGIT
           YRYPYLRVERERRRRHMEGVKDRDAG"
gene      5910. .7145
           /gene="MK1623"
           5910. .7145
           /gene="MK1623"
           /codon_start=1
           /transl_table=1
           /product="Uncharacterized protein specific for M.handleri,
           MK-39 family"
           /protein_id="AA02836.1"
           /db_xref="GI:1988835"
           /translation="MRSNTPPERGDSLTLMWLLMLPMPRAQADPTSHLYEHPT
           VPYEPESPLVAVVELLDEKKAELSLLELIGIPSEAAVGLADYVFGSGRGTR
           CLVYNPQDTVEVPEWEDRLPLRLDDHLLTELICLPVREGITLQPAVLG
           WNLMOGCTKAVRLHKGEDPTAVIAYVPEIETALAFESLLDPDNRVARAMERS
           SEVDAIDVYRDLRELVSRISGSDYDREYFESMHLIVSKIEELDEVFLPA
           LIEMCAIVRIVETSLITAFALYLPDNPANAVLERRAVVNSPYFOLLRLMPLRSTY
           AMKSVLCDAVTVADSLIEPAAKPVLSALRLDSESSRTTLEETTSAYAKNPLEALGPEL
           RTVLELIDLTLMGLAMAC"
Query Match      12.4% Score 126.6; DB 1; Length 10685;
Best Local Similarity 48.4%; Pred. No. 1.2e-08;
Matches 419; Conservative 0; Mismatches 434; Indels 12; Gaps 2;

```

```

QY 140 TCGCCGTGATCGAGGCGAATACCTGCGCGCGAAGACGGAATCTTCATCTTGACA 199
    || || || || || || || || || || || || || || || || || || || || ||
Db 10546 TCATCTACCTGAACGCGAAGCTGTCCTCCCGGGAAGAGGCTAAATCTCCGCTACGACC 10487
QY 200 CCGGATTCGGTCACTTCGCACTGACCTGACGCGCGCATGTATGGACGCAACATCT 259
    || || || || || || || || || || || || || || || || || || || || ||
Db 10486 ACGGGTTCCTGTAACGCGCGGCGTTCGAGGGGATTCGGCCCTACGATGGCCGGATAT 10427
QY 260 TCCGGCTCGGCGACCACTGACCGGCTTCGACGCGGCGCTCAAGCTCGCCTGGAGC 319
    || || || || || || || || || || || || || || || || || || || || ||
Db 10426 TCAAGTTGGACGACGCTGACAGACTGACAGCTCCCAAGGCGCATCATGTCTGAGA 10367
QY 320 CCGGATTCGGTCACTTCGCACTGACCTGACGCGCGCATGTATGGACGCAACATCT 379
    || || || || || || || || || || || || || || || || || || || || ||
Db 10366 TACCGATGACCAAGGAGATGAGAGCCCTATCTGAGACCGTACGAGGCGCAACGAGC 10307
QY 380 TGGCGCAATGCTGTAATCTGACCGTCAACCGGGGATACGGAAGCGCAAGGCGGAGA 439
    || || || || || || || || || || || || || || || || || || || || ||
Db 10306 TCCGGACGCGTACATCCGGGTGTCTCCGCGCGCAAGGTGACCTCGGTTGGAGC 10247
QY 440 AGGACCTGTCCAAAGCTACCCATCAGGTGTACATCTACCGCATCCGCTACTGTGGGCT 499
    || || || || || || || || || || || || || || || || || || || || ||
Db 10246 CGGGAAGTGTCCGGAACCAACGAGTGGTG-----ATCATCGCGGAGCGCATGGAGC 10196

```

```

QY 500 TCCCGCCCGCCGACGAGATCTTGCGCACACCGCATCTGCGCGGCATCTGCCCGCGC 559
    || || || || || || || || || || || || || || || || || || || || ||
Db 10195 CGTTGATCGGGGACCTGTACGAGAAAGGATGATGATACGCGCTCGTGGCTGTA 10136
QY 560 CCGGCGCAACACCGCGACCGGACCATCAAGAACTAACAGATGGGGTGTATCTCACCGCAG 619
    || || || || || || || || || || || || || || || || || || || || ||
Db 10135 TTCCACCGGACGCGCTGATCGGAGATCAATGCTGCAACTACCTGAAACACATCTCG 10076
QY 620 CCAGTTTCGAAAGCCAAAGACCGGTGTCGCGGACCGCGATCTGCTGACCTCGGACACT 679
    || || || || || || || || || || || || || || || || || || || || ||
Db 10075 CGAAGATCCAGACCAACTGCGGCGCGGACGACGATCATGTTGGACACAGAGGCT 10016
QY 680 GCGTGGCCGAAGGTCCGGCTTCAAGCTGTGATCTGCAAGAGCGGCAAGCTGCGCTCC 739
    || || || || || || || || || || || || || || || || || || || || ||
Db 10015 AGCTCTCGAGGGTACCGGTGACAGCTGTCTGTGTGAAGAGGAGCGGACCGTGTACACT 9956
QY 740 CGTCCCGGAACGGGTTGCC---GAGCATCACCCGTAAAGAGGTGTGCAACTGCGCGAC 796
    || || || || || || || || || || || || || || || || || || || || ||
Db 9955 CACCGAGAGACACTATTTCTCCGGGGATACACAGGGGACCGATGATGAGATCTGGAG 9896
QY 797 AGATGGCATCGAAGCCACCTGCGGACGTCACACAGCGGTGAATCTACAGCGGAGC 856
    || || || || || || || || || || || || || || || || || || || || ||
Db 9895 AGCTCGGTATACCGCTGTGAAGAGAGCGGATTAACACTGGGAGAACTGTACGCGCGGAG 9836
QY 857 AGTTGATGGCGGTTCACACCGCGGGCGGTCACACCGATCAACTCGCTGATGCGGAG 916
    || || || || || || || || || || || || || || || || || || || || ||
Db 9835 AGGTGTTCTTAACGCGGACGCGGAGGTTGCGCCCGCTTGAAGAGGTGACGAGCGGGA 9776
QY 917 CCGTGGCAACGCGGACCGCGGTCTCACTGACGGTGACATCCGGACCGGTTCTGGCGC 976
    || || || || || || || || || || || || || || || || || || || || ||
Db 9775 AGATCGCGGAGAGATGCGCCAGGACCATCAAGAGCGGTATCATGAGCGCTTCGCGGAGC 9716
QY 977 TGATGACGAGCGCGCGCGCTGAT 1001
    || || || || || || || || || || || || || || || || || || || || ||
Db 9715 TCACGAGAAGAGGATACGCGCGT 9691

```

RESULT 7  
AL646080  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AL646080 190050 bp DNA linear BCT 07-DEC-2001  
Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;  
segment 5/11.  
AL646080 AL646053  
AL646080.1 GI:17431113

Ralstonia solanacearum.  
Ralstonia solanacearum  
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
Ralstonia.  
1 (bases 1 to 190050)  
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,  
Ariat,M., Billault,A., Brothier,P., Camus,J.C., Catolico,L.,  
Chandler,M., Cholone,N., Claudel-Renard,C., Cunne,S., Demange,N.,  
Gaspin,C., Lave,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,  
Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,  
Weissenbach,J. and Boucher,C.A.  
Genome sequence of the plant pathogen Ralstonia solanacearum  
Unpublished  
2 (bases 1 to 190050)  
Boucher,C.A.  
Direct Submission  
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CRPH, 27 rue Juliette Dodu, 75010 Paris, France, INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,  
31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire  
INRA, BP27, 31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

FEATURES  
source

1.190050  
/organism="Ralstonia solanacearum"  
/strain="GM11000"  
/db\_xref="taxon:305"  
/plasmid="megaplasmid"  
51..362  
/gene="Rsp0643"  
/note="RS05579"  
51..362  
/gene="Rsp0643"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHEICAL PROTEIN"  
/protein\_id="CAD17794.1"  
/db\_xref="GI:17431114"  
/translation="MKTSPSPNPAAGPASPAGDGVAFQLEHRYHTTIPR  
RIGGGLMRAAFVSLDSRLVAVGTGVQAFRTPIGARNAVDAARSFIDARLKACRP  
A"  
397..555  
/gene="Rsp0644"  
/note="RS05578"  
397..555  
/gene="Rsp0644"  
/function="miscellaneous; unknown"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"  
/protein\_id="CAD17795.1"  
/db\_xref="GI:17431115"  
/translation="MERQQLFAIIVIAVSAVSLARDLNDERPFGQAALVGH  
ATASAAGR"  
complement(575..1135)  
/gene="Rsp0645"  
/note="RS05577"  
complement(575..1135)  
/gene="Rsp0645"  
/function="miscellaneous; not classified regulator"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD17796.1"  
/db\_xref="GI:17431116"  
/translation="MKRISTEPDVMKKVAIKFRSPAAPPOGAALSDORTRAIL  
KMLORDASISNALADKVSLSPPACLRERLERKEGCTRGVVALLEPRLDICTVMI  
GVVDRSPESPFADEKAVOKVPGCLCHVYGVGEEDYMLVTRNDSTYNRLHAEQL  
VLPVGRQITPTFKLVSTIOLPYA"  
1244..2260  
/gene="Rsp0646"  
/note="RS05576"  
1244..2260  
/gene="Rsp0646"  
/RC\_number="4.1.99.4"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : probable  
Gene name confidence : hypothetical

predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE  
DEAMINASE (ACC DEAMINASE) PROTEIN"  
/protein\_id="CAD17797.1"  
/db\_xref="GI:17431117"  
/translation="MNLNHPHPLTFEGPTPIQPLKLSAHLGKVELYAKRDNCSG  
LAFGKNTKLEYLVPEVLAVGYDTLVSIQSIQNTQVAAVAHLLKLVQVEM  
VAVDAVYDVRCNIELSRILGADVRLDAGPDIGIRPSMEQMEDVRAGGVPPIPA  
GSEHPILGICVGFVFAEEVROOEAEFGPEFDYIVYCSYTGSGFAGMGVPAADGADR  
VIGIDASAKPEQTRQILRIARDTAKIVELGIDPEDDVYDTRIGGPEYGLPNEGTL  
EALRICARQEGMLTPVYEGKSMHGHIDVRGGERPEGSRLVLAHLGVPALMAYSL  
FRNG"  
complement(2336..3508)  
/gene="Rsp0647"  
/note="RS05575"  
complement(2336..3508)  
/gene="Rsp0647"  
/RC\_number="4.2.1.17"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : putative  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"  
/protein\_id="CAD17798.1"  
/db\_xref="GI:17431118"  
/translation="WSALPKPAEPAAAEPEVLEHYVNVRAILTLNRPASLNLASHG  
VRLSALLAQCRDPRIAVAVLRAGCEKFCAGGDVRLRLAROGAAGPDGLOPE  
VDEYRLDYLALRFPKPVYALDGIAMGGMIGGAGRLRVLTESRTIAMETRGFLP  
DVGATHEFGAMPPEMALYIGLTGTLISGADLHGLADVCVPAEMLSLSPDRILRSH  
AGDPDLMQALRVPEPCNTVPAATARTPTVYRFDRESTDRIYATVRESLEND  
PAREARQIOLATIDAMKHSPTMLHVTREALLRQRLTLCGCFRMEIGVARIIEED  
FCBGVRAHLVKKDRPAGPAPATVAGVREVRHPLSSPWGAAHPLAHLGAOPOAA"  
complement(3505..4305)  
/gene="Rsp0648"  
/note="RS05574"  
complement(3505..4305)  
/gene="Rsp0648"  
/RC\_number="4.2.1.17"  
/function="small molecule metabolism; degradation; carbon  
compounds"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
Predicted by Codon\_usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"  
/protein\_id="CAD17799.1"  
/db\_xref="GI:17431119"  
/translation="MALDLWLDHGTVALTLTSSPPNAAFTEGLROLQAVETLADAP  
RVRLVITGDGPKFFSAGDLNFTASGDRDVARDEAARFGAFAEALONAPVYTAIN  
GYAMGGLCALACDILRIARHQLAVPETAVALCCGCGTOLPMLVGEKAKRML  
TGERIDQOTALRIGLVEEVDYDGAARPAALAMARRTYTSPRAVYTSKALIHQARGV  
PRTAALAVEREPRVDLFGHPDQREGVNAFLKKRPKMTNAREEALP"  
complement(4302..5198)  
/gene="RmsB"  
/note="Rsp0649; RS05573"  
complement(4302..5198)  
/gene="RmsB"  
/RC\_number="1.1.1.31"  
/function="small molecule metabolism; amino acid



biosynthesis; isoleucine/valine"  
 /note="Product confidence : probable  
 Gene name confidence : putative  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by Framed  
 /codon\_start=1  
 /evidence-not\_experimental  
 /transl\_table=11  
 /product="PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE  
 OXIDOREDUCTASE PROTEIN"  
 /protein\_id="CADI7800.1"  
 /db\_xref="GI:17431120"  
 /translation="MKIAPIGLGNMGAPMARNLKAGHALSVEDLDARAVOSIYVDA  
 SAAPSPRAATGACAITMLPAAHVHYLTADGDIAGIAKVPIDSDTIDPSAK  
 ALAAITAHNPNPVDAVPGSGGTGAAGITLFWAGAAVEEVRVTSGLMELVHC  
 GERAGAGAKICNNLVIIGITMAVAPAMSLGSLGIDPVYLTGIRVSGRMSSTPY  
 NPPGVYETAPSSRGYTGFGFDLMKLDGLADDAKSVQPYLGLALAQOLYQASSS  
 QGAGKIDFSAVIRLYRPAGAR"  
 complement(5220..6767)

Query Match 10.8%; Score 109.8; DB 1; Length 190050;  
 Best Local Similarity 47.7%; Pred. No. 1.2e-06;  
 Matches 398; Conservative 0; Mismatches 422; Indels 15; Gaps 2;

135 CGGCGTCCCTGATCGAGGCGGATACCTGCGCGGAGAACGAGAGATCTCATCTT 194  
 Db 82040 CGGATGCTCTTCGAGGCGCAATAGTCGCGCGCGAGCGCAACGCTCTCATCAT 82099  
 Oy 195 CGACACCGGATTCGGTTCATTCGATCTGACTACACGTCGCGCATGTATGGCGGCA 254  
 Db 82100 GACACCGCGGCTACCTTGGCGATGCGATACGAGTGCACCGCGTACGACCGCGCAA 82159  
 Oy 255 CATCTTCGCGCTCGCGCGACCTGACGCGGCTGCTCGACGGGCGCTCAAGTGCCT 314  
 Db 82160 GCTGTCACACACGACGCGCCACCTCGCGCTGACGCGCTCGCTCGCGATCGCGAT 82219  
 Oy 315 GACGCGCGGATACACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 374  
 Db 82220 CGACACCGGATTCGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGAT 82279  
 Oy 375 GACGCGCGGATTCGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGGATGCGG 434  
 Db 82280 CGGCGTGAAGAGGCGGCTGCTGATCATCGCTGCGCGCGCGCGCGCGCGCGCTT 82339  
 Oy 435 CGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 494  
 Db 82340 CGGCGATTCGCGCGGAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 82399  
 Oy 495 GCGCTTCG 554  
 Db 82400 CGACACCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCT 82453  
 Oy 555 CGGCG 614  
 Db 82454 CTGGAAGCGCTCGGACATCAAGCGCTGCGGCTGCGGCGGAGGATCGCGCGAGCT 82513  
 Oy 615 CGCAGCGAGTTTGAAGCGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 674  
 Db 82514 GCG 82564  
 Oy 675 CAAGTTCG 734  
 Db 82565 GACCGAGGCGCGCTGCTGACACGCGCTTCAATCAACACCGCAAGGCGCTGATCAGCG 82624  
 Oy 735 CTCGCCGCTCG 794  
 Db 82625 TCCGCGTGTGATGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82684  
 Oy 795 CGAGATGGGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 854  
 Db 82685 CGAGGAGGCGCGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 82744

Oy 855 CGAGTTGATGGCGGTCACACCGCGCGGCGGTCACACCGATCACTCGATGCGCA 914  
 Db 82745 CGAGGCGTCTTCTTCACACGAGGCTTCACCTTCGATGCGCGCGCGCGCGCGCGCG 82804  
 Oy 915 GCGCGTGGCGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969  
 Db 82805 ACCGTTGGCGGCG 82859

RESULT 8  
 AE005101/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
 Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Shrogha,J.,  
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,  
 Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,  
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angewine,C.M., Dale,H.,  
 Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,  
 Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
 Dassarum,S.  
 Genome sequence of Halobacterium species NRC-1  
 Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)

TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS

2 (bases 1 to 12730)  
 11016950  
 20504483  
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,  
 Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Shrogha,J.,  
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welti,R., Goo,Y.A.,  
 Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,  
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angewine,C.M., Dale,H.,  
 Isenbarger,T.A., Peck,R.F., Pohlischrod,M., Spudich,J.L.,  
 Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,  
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and  
 Dassarum,S.  
 Direct Submission  
 Submitted (14-JUL-2000) Institute for Systems Biology, 4225  
 Roosevelt Way NE, Seattle, WA 98105, USA  
 Location/Qualifiers

FEATURES  
 source

gene  
 CDS

1..12730  
 /organism="Halobacterium sp. NRC-1"  
 /strain="NRC-1"  
 /db\_xref="taxon:64091"  
 complement(70..597)  
 /gene="pyrE2"  
 /note="VNG21186"  
 complement(70..597)  
 /gene="pyrE2"  
 /note="pyrE2"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orate phosphoribosyl transferase"  
 /protein\_id="AG20258.1"  
 /db\_xref="GI:10581533"  
 /translation="MSATDDIVSALPAADAVOFGELSHGTSSEYVYKLYFEDPE  
 CISAIAAFADRIDETTLGVALGVPPLAAVATFAGVPIYARQAKETGIANRIE  
 ERALDEEVVVEDIATITGOSAVDAVDALRDAGATVNRALIVYDREGRELLAEGV  
 complement(653..1297)  
 /gene="VNG2119C"  
 complement(653..1297)  
 /note="VNG2119C"  
 /gene="conserved hypothetical protein"  
 /codon\_start=1

gene  
CDS  
/transl\_table=11  
/product-"Vng2119c"  
/protein\_id-"AAG20259.1"  
/db\_xref="GI:10581534"  
/translation="MGIIPTSRSPHYLNRPIGIRGFLPGARPPTRMOLVGVVAV  
MAMPAVNNNAVLAGGPRIDGSGSLGRRLLGGKTRGTAAGVALLAVLN  
ALRPAADALGVLPAPFPRAKGTLEAGAVGIIASFLKRTGRORGAAPVVDLD  
FVVVALTLALAVPAWVDTFGPLVTVAVLPLHLTLNGIAYALGVDFEW"  
1284. .2123  
/gene="yusm"  
/note="VNG2120g"  
1284. .2123  
/gene="yusm"  
/note="yusm"  
/codon\_start=1  
/transl\_table=11  
/product-"proline dehydrogenase"  
/protein\_id-"AAG20260.1"  
/db\_xref="GI:10581535"  
/translation="MIPPIANRFVAGETPAEIAEYADINDRGVTAIINLGEHYHER  
PPADEDAARLRVLADVGDSDVACVSPQIGLDVGEPEFENLETIAAABHDV  
PAWMDDEHTTADTDESLARTAGNGLCVQANLKRTEODLDRLDVGVKRLV  
KGAIVEDPDVAYODKPDVNEAYRTLEFPERBERDGVAVGSHDPAMIISYAALHREYV  
TDEYVGMKGIREDAGTLELAIGVPTIYQVYPIGKNFSIFIRRAEKKNLVALRAV  
LGR"  
2195. .2812  
/gene="VNG2121c"  
2195. .2812  
/gene="VNG2121c"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product-"Vng2121c"  
/protein\_id-"AAG20261.1"  
/db\_xref="GI:10581536"  
/translation="MGSKRDFASGLIVPIVYLVIYVYLMFGLISNTLTLELDT  
OQAVATLAVFLVLYFSVGYLMRTAGSILFALIDVYVNRPLPYARIYNASKAVET  
VLSDGGEFOQPRVYPPWPMRLTAKTKKAPDGEVEVFETAPITTTGFEVEPE  
DLSEDEGVEDALTRVLSAGFGENDEVAIDELLPEDDENE"  
complement(2850. .3776)  
/gene="11ve2"  
/note="VNG2122g"  
complement(2850. .3776)  
/gene="11ve2"  
/note="11ve2"  
/codon\_start=1  
/transl\_table=11  
/product-"branched-chain amino acid aminotransferase"  
/protein\_id-"AAG20262.1"  
/db\_xref="GI:10581537"  
/translation="MPSEEMVDTITMDGEFVMDENDAQVHVLTHALHYSGIFEGVR  
VYTDNGPAIFPMDDHLOREFESAPRYDIDFETDELDTAVELIQRDLOGCYTRP  
LAYVYESLGVRGDCPDVVAIAAMPWGLVGDALLENQVDMVSTMRKHAASQIPTN  
AKTGLVNSMLAGEARNGFTETAILNKEGVNAEGENFELVRDEGLPGISGS  
IIDGTRDVTILAEERGVTHDVVISRKGELHTADELFTGSALEVPIKQVDNVEI  
GSGTRPVTEELQTFPDLVEAGDEEMFHY"  
3908. .5383  
/gene="nhac2"  
/note="VNG2123g"  
3908. .5383  
/gene="nhac2"  
/note="nhac2"  
/codon\_start=1  
/transl\_table=11  
/product-"Na+/H+ antiporter"  
/protein\_id-"AAG20263.1"  
/db\_xref="GI:10581538"  
/translation="MSDEPLAYEDIDARKPTLEALVPLVYVYVSGVSGYLGLA  
PRAPIKSTTTPGLFARYHLGIDMERVSADALAGKGLQAILILEVYGLIATWISA  
GITPLIMYTGALTPAVFLPVTLAAVAVFAGISSWTGTLGAVGIGSGIGIP  
LPMTAGIYSGVAGDKQSPISDTNLAAYTNADLVIHNGMRIGTVAIAGLSIYAV  
AVIGLIVDGANANTAIISGLAGSYALGPLVETPLVYFGLAFRGYPALSIYAV  
AGALTIVAGASFTAAWDIFLNGTTPPTGSELVNNLITGGIAGSANTIVVVAALS

gene  
CDS  
/transl\_table=11  
/product-"Vng2124c"  
/protein\_id-"AAG20264.1"  
/db\_xref="GI:10581539"  
/translation="MASYGVALLVPLLAIVLMTTROYLSLFGVAGALVAVOAN  
PIATALMDMLVAVRPPNKKFVILIFMGACAPTRSGIILAEEMRISLTA  
RESOTIMLIGVETIFEDSYTSTVYGNATRELADDSREMAHYVLDSTIAPVITG  
PVSNTIGVSKITLIGLEAANVSAELETTPGFLDSIPNITICMAFPVWGVTIT  
QRFEGPMLDAEMRAATGTDITADGATPLSDVNDVGPSTANPILNFPGLVAVLAV  
GLGSMWLYGGYKPRVDIATLOETDVMGILHQAFAFMVSGFVGAALGTMDEGAT  
DTIVGFEKTNMIALIIVLANGIGMAEKVTADFIETLVGSGVAPGLPLVEFLA  
MVAATTTGSMGTALILPPVAIPRGINAGAAVPLVGVFGAIGCDHSSPISDT  
VMSSTFASGSHIDHYTQIPRAATACVTIMLVLYAVGLRSPVALPLATVLTIGAV  
VVLKLDTRKRGLEPVRADAVDADREAAANGTSLDPAATSYDLSLVPAAVGI  
VLLYLSLVEFPAALG"  
complement(7210. .7716)  
/gene="r1ba"  
/note="VNG2125g"  
complement(7210. .7716)  
/gene="r1ba"  
/note="r1ba"  
/codon\_start=1  
/transl\_table=11  
/product-"GTP cyclohydrolase II"  
/protein\_id-"AAG20265.1"  
/db\_xref="GI:10581540"  
/translation="MRTDAGSLICVALGDAVADFALPFTADELHPADHGLGYDD  
RSSFLVYNHRDITVGTDDRALTIELQAAADPAADVPEPTFRAPGHVHLRAP  
SLSAROGTEELGIALAGAGVPAVVVCEMLDADTGARSVADAKRYASRHDIPYEG  
SDLVAAFO"  
complement(7865. .8593)  
/gene="VNG2126c"  
complement(7865. .8593)  
/gene="VNG2126c"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product-"Vng2126c"  
/protein\_id-"AAG20266.1"  
/db\_xref="GI:10581541"  
/translation="MRGACQPPANSATSTGVGYDELAVALKLALDGAHGEVYKSC  
GDLASRLDASSQTSRLQALDDADHYTRDLYSDGQWITVTDAGRHLAKHEYDRI  
FEDEGELALVYSGMEGRNYTSLPGYRORFAEKIGYPRPGTINVDLPDGOBAR  
AGIOLAGVIDAMEDDEDRYGSATCYACTVVDGDPFGAHYIIPDRTHDDOLEI  
IAPKLERGLDDEDTTRVEA"  
complement(8630. .9607)  
/gene="VNG2128c"  
complement(8630. .9607)  
/gene="VNG2128c"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product-"Vng2128c"  
/protein\_id-"AAG20267.1"  
/db\_xref="GI:10581542"  
/translation="MSSEDPKQVDDPVDYHNHNHTAATQTCGTANMARBGCTYKNAF

Query Match 9.48; Score 95.8; DB 1; Length 12730;  
Best Local Similarity 47.7%; Pred. No. 0.00019;  
Matches 428; Conservative 0; Mismatches 442; Indels 27; Gaps 4;  
DB 3775 TGCCAGTTTCAGTGATGAGCTGCACACCACTGTGATGAGCGGAGTTCGTGACT 3716

170 CCGAAGAGGAGATCTCCATCTGACACCGGATTCGATCCGATCTGACCTACA 229  
 3715 GGGAGGAGCAGACAGTACAGTCTCTACCCAGCAGTCCAGTCCGAGCTTTCG 3656  
 230 CCGTGGGCGATGATAGGCGAGAAC-----ATCTCGGCTCGGCGACACC 277  
 3655 AGGGGGTCCGCTTACGACACAGAACGGCCCGCGCTCTCCGCTGGAGCCACC 3596  
 278 TGGACCGCTTCTGACAGGGGCGTCCAAAGTGGCCCTGGAGCGGCGGTACAGCAAGGAGC 337  
 3595 TCCAGCGCTTACGATGAGGCGAAACCTTACGACCTCGATCGAGTTCAGCCCGGAGC 3536  
 338 AATGGCGGAGATCACCAGAAAGTGTGATGTCGAGAGTGCAGGATCGTTTGTGA 397  
 3555 AATCACCAGAGCCAGCTGAGAGTCTATCCAGGCGCAAGACCTCGGCTCTGTTATATCC 3476  
 398 ATCTGACCGTACCCGGGATACGGAAGCGCAAGGCGCAAGGACCTGTCCAGCTCA 457  
 3475 GCCCGCTGGCTACTACGAGGCTACGAAATCCTCGCGCTCTC-----CCCCGGGACTGCG 3422  
 458 CCGATCAGGTGATATCTAGCCCATCCCGTACCTGAGGCGCTTCCGCGCGCGAGCAGA 517  
 3421 CGACCGAGTGCATCGCGCTGGCGCGGCGGCGGCTACCTCGGCGAGACGCCCTCG 3362  
 518 TCTTGGGACACCGGCGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577  
 3361 AAAAGCGGCTGAGTGTATGTGTCGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3302  
 578 ACCGACCATCAAGATCACTAGTGGGCTATCTCACCAGCGCATTTTGAAGCCCAAG 637  
 3301 CGAAGCGCAAAACACCGGCTCTTACGCTCAACAGCAGTGTGCGGCGGAGAGCCCGG 3242  
 638 ACCGAGTGGCGGACCGGCGATCTGCTGATCGAGTGTGCGGCGGCGGCGGCGGCGG 697  
 3241 GCAAGGCTTACCGAAGCCATCTCTCTCAACAAAGGCGCAAGTGGCGGAGGCCCG 3182  
 698 GCTTCAAGTGTGATCTCAAGAGCGCAAGCT-----GGCTCCCGCTCCCGGAGAG 751  
 3181 GCGAAGACATCTCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3122  
 752 CGTTCGCGGCGGATCAACCGGTAAGAGCGTTCGAACTGGGCGGCGGCGGCGGCGG 811  
 3121 TCCTGAGCGGATCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3062  
 812 CCACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 868  
 3061 TCACAGAGAGAGTCTCGATCTCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGG 3002  
 869 TCACAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 928  
 3001 CCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2942  
 929 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 985  
 2941 GAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2885

RESULT 9  
 AP003000/c 348077 bp DNA linear BCT 15-MAY-2001  
 DEFINITION Mesorhizobium loti DNA, complete genome, section 7/21.  
 ACCESSION AP003000  
 VERSION AP003000.2 GI:14023048  
 KEYWORDS  
 SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.  
 ORGANISM Mesorhizobium loti  
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Phyllobacteriaceae; Mesorhizobium.

REFERENCE  
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Ideana, K., Ishikawa, K., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,

TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 JOURNAL  
 COMMENT  
 FEATURES  
 SOURCE

Mochizuki, Y., Nakayama, S., Nakazaki, N., Shlimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.  
 Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti  
 DNA Res. 7 (6), 331-338 (2000)  
 21082930  
 2 (bases 1 to 348077)  
 Kaneko, T.  
 Direct Submission  
 Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: kaneokkazusa.or.jp)  
 URL: http://www.kazusa.or.jp/rhizobase/  
 Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934  
 On May 11, 2001 this sequence version replaced gi:11994975.  
 Location/Qualifiers

1. 348077  
 /organism="Mesorhizobium loti"  
 /strain="MAFF303099"  
 /db\_xref="taxon:381"  
 /complement(169..1065)  
 /gene="ml12551"  
 /complement(169..1065)  
 /gene="ml12551"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional regulator"  
 /protein\_id="BAB49656.1"  
 /db\_xref="GI:14023049"  
 /translation="MALDMDKLRVFFHAAAGSFTHAATLHLSQASISROVSALEHD VGPLFRNRHROEFTLEKPEISQILIANEIDLTWRQDCAIRLROPOOLIORLP LGAGMTERHROEFTLEKPEISQILIANEIDLTWRQDCAIRLROPOOLIORLP VVPHLYAAFSYVAKHGKPAVSSELNRIHVFGLPVSHSELMLETVDGDFEGOR VPSIQINDILSIRAVOGAGIAMPDYVINRDSGLVOLLPTEVPSEDTYFAYPDAM KNOAKLHVFEDFTIARASWSF"  
 /complement(1156..2130)  
 /gene="ml12552"  
 /complement(1156..2130)  
 /gene="ml12552"  
 /codon\_start=1  
 /transl\_table=11  
 /product="thioredoxin reductase"  
 /protein\_id="BAB49657.1"  
 /db\_xref="GI:14023050"  
 /translation="MTTKHAPVLIISGRAGTYAAYARAKLPMYVAGLQGGQGM ITVDENVYPEPADP ICGPMLMEQMMQAEHVGDIIINDITTEVDLNVRFKAGSGT- TYRADALIATGAQAKWLGIPTEODPMGVSACATCDGFEYRGDVAAYGGNSAVE EALYLSNLKSVYVYIHRSDPFAERILIRELQKDNVRYINDVDEITGRGKAPLP PSVEGELKRAVYGAETHLKVDDVFVAIGHAPVELFVGLKOKPNGYLMTPNSTRF DNRGVFAAGVTDVYKRAVYTAAGLQCMALREKYLAGEVHREAAE"  
 2378..2857  
 /gene="mlr2554"  
 2378..2857  
 /gene="mlr2554"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable transcriptional regulator"  
 /protein\_id="BAB49658.1"  
 /db\_xref="GI:14023051"  
 /translation="MPLKADLADIMKILNELQADGRMTNVELSNRVSISAPPCLRV KLEFAGILINGYRLNALPALGDVAVFCVGHQADDELFTFAEFTGMPGIVDAM MVSGESDFLHCLASDLGAFQTVIEELSTPVPVIVRALTIRVKBGIVSFERS"  
 3093..3887  
 /gene="mlr2557"  
 3093..3887  
 /gene="mlr2557"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glucosyl transferase"  
 /protein\_id="BAB49659.1"  
 /db\_xref="GI:14023052"  
 /translation="MTRFASPVSLIICVNEADMIGPCLESIDFCAEIIIVDSGSTDG



QY	553	CGCCGCCGCCGCCGACAACCGTGGACCCGACATCAAGAACTACCACTGGGGGATGATCTC	612
Db	78639	CGGCCCGACCCCGACGGCGCCCTCGAATTCGAAGGCGCGCGGCTTCTACATGATCTGC	78640
OY	613	ACCGCAGCAGTTTCGAAAGCCCAAGACCGCTGGTGGCGCCACCGCGATCCTGCTCAGTCG	672
Db	78639	ACCATGTCAAAGACACACGCCCGCCGAAGCAAAAGGCTACGGCCGACCATATGCTGCACTCGG	78580
OY	673	GACAACTGCTGGCCGGAAGTCCGGGCTTCAAAGTGTGCATCGTCAAGAGCGGCAAG---	729
Db	78579	CGCGGCCAGGTGCGCGGAAGCAACCGGCGCCAAACATCTTCTTCGTCAAGACGCGCAAGATC	78520
OY	730	CTGCGCTCCCGCTCCCGGAACGCGTTGCCGGGCAATCACCCGTGAAGACGCTGTTCGAATCG	789
Db	78519	CATACGCCCAAGCCCGATGTTGTTCTCTCGACGCGCATCACCCCGCGGCGGATCTCGGGTTC	78460
OY	790	GCCGACAGATGGCGATGGAAGCCACCCCTGGGCGGACGTCACCAAGCCGTAATCTACGAC	849
Db	78459	GCCAGGGATCGTGGGCTTGGAAGTGAATCAACCGCGCATCTCGCCGAAAGATCTGAAGGC	78400
OY	850	GCCGACGAGTTGATGGCGTCAACACCGCGGGCGGGCTCACACCGAT	896
Db	78399	TTTCGAACAATGCTTCTTGACCGGAACCGCGGGGGAAGTAACCCGGGT	78353

gene  
CDS

```

/note="Product confidence: hypothetical
Gene name confidence: hypothetical
Predicted by Codon-usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD14086.1"
/db_xref="GI:17427567"
/translation="MALTSSAFNFALGSIILNTNVLSGNAVVALAARNLPRROQRQ
AIFMSSAAIILRLRIADIVMSIDNVVAAVAAAEKGPCTDVALLVIGLISEADEDAEHP
HDGIMPAIRTLRLINDIVMSIDNVVAAVAAAEKGPCTDVALLVIGLISGSIPLIFGSS
LTVLALMRPLPIVTVLAAALGLVLAGMLVTPDPDAGWEMQALPYADVAIGCLGALLIV
IVGMWGRSLDQA"
951..1472
/gene="p1a1"
/note="RSC0557; RS04908"
951..1472
/gene="p1a1"

```

RESULT 10	AL646060/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
	AL646060	212050 bp	DNA	linear	BCT 07-DEC-2001			
	Raistonia solanacearum	GMI1000 chromosome,	complete sequence;					
	segment 4/19.							
	AL646060	AL646052						
	AL646060.1	GI:1427566						
	Raistonia solanacearum.							
	Raistonia solanacearum							

gene  
CDS

```

/function="structural elements; cell exterior; surface
structures"
/note="product confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE TYPE 4 FIMBRIAL PILIN SIGNAL PEPTIDE
PROTEIN"
/protein_id="CADI4087.1"
/db_xref="GI:17427568"
/translation="MASHSHSNRRRGQRTLLIEMIVATLIGILAALAIPIRYDIT
SRVVEAGALIGQAKALVEALMAAQSISLGASLPARKNSALINDSTGELIIVV
APAAVEAGANTLIVLTPYSGSSSVFGNLQAGSTPPAVLVVCAAGSKSMVNTYQRT
ALIPKRYASECR"
1739..2245
/gene="piliA"
/note="RS0558; RS04907"
1739..2245

```

REFERENCE AUTHORS	TITLE
1 (bases 1 to 212050) Salanoubat, M., Genin, S., Attiguenave, F., Gouzy, J., Mangenot, S., Ariat, M., Billault, A., Brotlier, P., Camus, J. C., Catolico, L., Chandler, M., Choise, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saudin, W., Schlex, T., Siquier, P., Thebaud, P., Whalen, M., Winkler, P., Levy, M., Weissensbach, J. and Boucher, C. A.	Genome sequence of the plant pathogen <i>Ralstonia solanacearum</i> unpublished
2 (bases 1 to 212050) Boucher, C. A.	Direct Submission
Submitted (05-DEC-2001)	Genoscope and CNRS UMR-8030, 2 rue Gaston

EMEREAUX, C. 1987. Les interactions Moléculaires des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CERH, 27 rue Juliette Dodu, 75010 Paris, France, LMCM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA, 2 rue Gaston Cremlieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex

Christian.Bouchere@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

Location/Qualifiers

I. .212050

FEATURES

source

gene  
CDS

```

/evidence-not_experimental
/ransl_table=11
/product="TYPE 4 PIMRIN-L P11N SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD1088.1"
/db_xref="GI:17427569"
/ransl_table="KMSNRHLKRVKQGFLLIELMIVAIVGILAAIPAYQDYTVV
AVTELSLTAQAKLVAENNAQNSDLSVGSSVPEPTKRVNVLPIAGQVYVQGIT
VTTTAGGGTALVYTAGTALPVSSAPSGPIMMTCYAKQKAQASSVAPSSMTSLAL
KPAAPBCR"
2351. 4060
/gene="RSC0559"
/note="RS04906"
2351. 4060
/gene="RSC0559"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/ransl_table=11
/product="HYPOTHEMETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD1089.1"

```

FEATURES	
source	Location/Qualifiers
gene	1..212050 /organism="Ralstonia solanacearum" /strain="GMI1000" /db_xref="taxon:305" 59..757 /gene="RSC0556" /note="RS04909" 59..757 /gene="RSC056" /function="miscellaneous; hypothetical/global homology"
CDS	

```

/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEMETICAL_TRANSMEMBRANE_PROTEIN"
/protein_id="CAD14089.1"

```







```

/transl_table=11
/product="Lipopolysaccharide core biosynthesis protein
Maap"
/protein_id="AAG08394.1"
/db_xref="GI:9951295"
/translation="MRLVLEPPKRLMNGDPPEAYHALQKYYRELEGRTLRTEGD
GRGFTVILHKGICMGEIATKULTRKLPVLCARQEWQKIRLHAGVATMVAAYGEG
SDPARQHSFVTEIELAVDLEVEFSDMRPPRLKRLALVEAVAMGDMRAGVN
HRDQYICHLPLHTDKPVASDPRFLSVDLHRAOTRDATPKRMKDLAALYFSLDIG
LTRDKLRFRTYFRRLREILRDENGLLAWMERKAEKLYERQYRDLL"
complement(10200..11321)
/gene="waag"
/feature="PA5010"
complement(10200..11321)
/gene="waag"
/codon_start=1
/transl_table=11
/product="UDP-glucose:(heptosyl) LPS alpha
1,3-glucosyltransferase Waag"
/protein_id="AAG08395.1"
/db_xref="GI:9951296"
/translation="MTLAFILYKYPFGGQHPDPMRLALECCORGHDIRVLTIMED
VPOGEVLVAPRSIFNHRNEKETAMRADLRPVORVIGPKMGIDVYAAADAC
FEERKQTLRLPLRQMGVRYRAGYBRAVDPASKETIIMISVQDPLVYKHGTAE
RPHILPPGISQDRAPANADVAERREFEEDLLLVQISGFKTGLDSLKAL
SALPKALRRTRRLIAGDDPKPFLQLIALGLNDVQILKRSIDIPRLGLADLLIH
Query Match 8.4%: Score 85.8; DB 1; Length 1756;
Best Local Similarity 48.2%; Pred. No. 0.0041;
Matches 336; Conservative 0; Mismatches 352; Indels 9; Gaps 3;

Db 13655 GAGAGGCTCTTCTCACTGGCACTGCCGGGGAAGTACGCCCATCGGAGTGCAGGT 13596
Oy 913 GAGCGCTGGGCAAGCGCGGAGCCGGTCCACGTGAGG 949
Db 13595 CGGAAGTGGCGCCCGCCGCGCGTGGCGGTGACCG 13559

RESULT 12
AE011153
LOCUS
DEFINITION
Methanosarcina acetivorans str. C2A, section 498 of 534 of the
complete genome.
ACCESSION
AE011153 AE010299
VERSION
AE011153.1 GI:19918472
KEYWORDS
Methanosarcina acetivorans C2A.
SOURCE
Methanosarcina acetivorans C2A.
ORGANISM
Methanosarcina acetivorans C2A.
REFERENCE
Galagan,J.E., Nusbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P.,
Fitzhugh,W., Calvo,S., Engels,R., Smirnov,S., Atencio,D., Brown,A.,
Allen,N., Naylor,J., Stange-Thomann,N., DeArrellano,K., Johnson,R.,
Lincoln,L., McEwan,P., McKernan,K., Talamas,J., Pittrelli,A., Ye,W.,
Zimmer,A., Barber,R.D., Cann,I., Graham,D.E., Grahame,D.A.,
Guss,A., Hedderich,R., Ingram-Smith,C., Kuetner,C.H.,
Krzyczk,I.J., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
White,R.H., de Macario,E.C., Ferry,J.G., Jarrell,K.F., Jing,H.,
Macario,A.J.L., Paulsen,I., Pitchett,M., Sowers,K.R.,
Swanson,R.V., Zinder,S.H., Lander,S.H., Metcalf,W.W. and Birren,B.
The Genome of M. acetivorans Reveals Extensive Metabolic and
Physiological Diversity
Genome Res. 12 (4), 532-542 (2002)
JOURNAL
MEDLINE
21929760
PUBMED
11932238
2 (bases 1 to 10993)
REFERENCE
Birren,B.
AUTHORS
Direct Submission
JOURNAL
Submitted (20-MAR-2002) Center for Genome Research, Whitehead
Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES
source
location/Qualifiers
1..10993
/organism="Methanosarcina acetivorans C2A"
/strain="C2A"
/db_xref="taxon:188937"
78..629
/gene="MA4346"
78..629
/gene="MA4346"
/codon_start=1
/transl_table=11
/product="hypothetical protein (multi-domain)"
/protein_id="AA07688.1"
/db_xref="GI:19918473"
/translation="MQENPEEKKREPSDFKAPGLETGYEKSWKPEKTEMEKE
AGIEPTPYRKESDFKAPGLETGYEKSWKPEKTEMEKEAGRETEEGKEG
EEERTYTKRKRLTKSRSDRLFEVCGGLKRYGIDPTFRLFLVLLAFVNGGIYLV
IILALIMPSESVOMPETRKSS"
complement(716..1255)
/gene="meng"
/feature="WA4347"
complement(716..1255)
/gene="meng"
/codon_start=1
/transl_table=11
/product="menaquinone biosynthesis methyltransferase
(2-heptaprenyl-1,4-naphthoquinone methyltransferase)"
/protein_id="AA07690.1"
/db_xref="GI:19918474"
/translation="MGRFLSDPFRKLOSPDKLIDRSIGEGHMYLEVYGGSAFTTF
VARVIGKEGYALDIOPMILMQKELSLPERNDRLNLIKLDGHNLPEDNSDL
VRLTIVQELDPKKNVLEKIRKVLKPGILAVTEFLPPDPYPLKSTTIRLGEAGLIL

```



gene  
1795..2088  
/gene="MA4348"  
1795..2088  
/gene="MA4348"  
/codon\_start=1  
/transl\_table=11  
/product="predicted protein"  
/protein\_id="AA07691.1"  
/db\_xref="GI:19918475"  
/translation="MEISGROCGGCTGTHPPPELKSIAKAPATLEELSMHRYGDES  
KEALQEDERKDLHWGSCSPYTAAGSFICRLMTASSADLCRVACFLFNK"  
3071..3955  
/gene="MA4349"  
/note="MA4349"  
3071..3955  
/gene="MA4349"  
/codon\_start=1  
/transl\_table=11  
/product="branched chain amino acid aminotransferase"  
/protein\_id="AA07692.1"  
/db\_xref="GI:19918476"  
/translation="MQMSDLITLNKRFYKDEAKVSYDHPITGDSVEEGIRAYNG  
RVRKLEHVDRLYSAKATAMDIPTREKTEITILEALKNNKADYIRPIYSRIGD  
LGIDPRKCGSPYVIAQSGMAYGDLVEGLGVSVCVRNAPDLSLNTNLTN  
NIIAKIEANKEGDEAIFLDHNGFCGSGDIFVKKDRLPFIISMLKQITRATA  
IELIDEGYKVISNGLFDLYADEIFVTGTAESAAPTILDRVIGCKPGPLMK  
MVEAFKIQSTQTOIYK"  
4274..6214  
/gene="moea"  
/note="MA4350"  
4274..6214  
/gene="moea"  
/codon\_start=1  
/transl\_table=11  
/product="molybdenum cofactor biosynthesis protein"  
/protein\_id="AA07693.1"  
/db\_xref="GI:19918477"  
/translation="MKRKEFRELVPVEANRIINRLQIRPEKENSALENAYKTLAED  
IYIEINPPSRATMDGYAARADTYACSETPEVYKLLGNIIPAGSDVSEKSGAGIV  
EIANGAPILPGADAVVVEYSENGVILFEPVAGGENYKAGSDILKFGVRLRGR  
KICREITGVLASIGKEKVPYRLPVGLISNGELVPGNLAKGKYDMVRITDEYET  
QEGASFLYGVVDEDTGIRKALETRAVSCAILNLSGSSSGDVMNRITDEYET  
LAGINIKPKPVPVIGIKVPIIGLPGNSALMIFNEFVAFPLKALGAEQVART  
ETGILGTVLSEKROOILPVGLVGRVYPADRSGLATLSLSEDFLEIPETIEEA  
GTAEVTLFEVEKPDLLIAGFCGPGLDVEDLSGFRFLTYGSGGFSALAGTAD  
IAGVNMSPKRNRESLPGELIVRNATINMGLSEVLYGYRREMGILVODSPVS  
GLEALPKRLINNRSGSTRALLEIKTEELAEKKSGKEFTSDIPGYGAKSEVYS  
CEAVLSKVDAGVGLRNCARNLKPYKFAEBEYDPLIREVYDVEVREKLTLSA  
EFMSKLPAQLQVTERGEITISY"  
7373..8467  
/gene="MA4351"  
7373..8467  
/gene="MA4351"  
/codon\_start=1  
/transl\_table=11  
/product="GTP-binding protein"  
/protein\_id="AA07694.1"  
/db\_xref="GI:19918478"  
/translation="MSSIOEIOIEVEDEIRKTOYNKATSHHIGRLAKIARMDEIEK  
KASSKGGEGVSKSGDGVTVLVGSPVSKSTLNNKVGAKSEVGEYETTLVVG  
VLEHKATIDFDPGLYKASGSRGKGVISYIRNSDPVITLLVYFOKKHVEYMD  
ELYVAGIRVDEVPDVTIKRDRGELINSTIDLDDEETIKAVLDREKIRHNAVILR  
ENITVDOLIVLGNRSYSLVYVNVVDLAYPOLIECKRLPKSIFISAHGSLINE  
NIKDSIYDLGFIYVLYLPOGOPADMEPLIVSGNIQICDRLHDFRRKRYAYV  
WGPSAKHGGVGLHEIMEDIIMTIYK"  
complement(8761..10872)  
/gene="MA4352"  
complement(8761..10872)  
/gene="MA4352"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein (multi-domain)"

BASE COUNT 3241 a 2382 c 2591 g 2779 t  
ORIGIN  
Query Match 8.2%; Score 84; DB 1; Length 10993;  
Best Local Similarity 45.9%; Pred. No. 0.0081;  
Matches 371; Conservative 0; Mismatches 425; Indels 12; Gaps 2;  
/protein\_id="AA07695.1"  
/db\_xref="GI:19918479"  
/translation="MTQNTDSDNPRRRKAKGTVOYKTLGVPVDLEPHYRDMQFLFN  
SLYLTKDADLLDKMEVTRKEADLLVSLIGLDEPMVLDLQCGRHLLELARQFSNV  
EGYDRSOYLIRKARPAKREMLQVRFPRGDRKLPYSDPFSVYTLIGNSGYPDSL  
HDRVLEVEFVVLKSGRVFTDAVDGDDKNNPDRSMEDRYEYFCRERALSDD  
RLICREVICRNDKGTIADQFYAEHLNRESLPELLTASGTSIPFPHFTPSVSGYD  
AGMGORILLSATVEKAMPSSLSLQNAESKLPNNVYVLDLPRKEDQVAPACVDD  
DLETVNRMKQALSEIPYKFTFLDRHETLLDLKRAKTDVLNLCDGFNNDPIKE  
LHVALLEOFNIPYTAGPQCSFCYDLSLVRAKREMLPVAKVAVTGDSDVSFS  
LSPLLIVPNSGDSYSYITOKSIYHSREIYDIKMERIKGADKPELLEFLPKDI  
SVGIGNPPCTVLPTEEDYSAPEDLPICGYEAKMLPSPYWKISKVAPGPEPT  
KEIYRCIALFTRLGCDYRKPDRDLDAERPRILLENPNPNCNGLAKMAAYN  
ISYSGMLAILLEAKKRCIGTSVKTIELQRAESRSPKSPDEYAAEEVEEAKGPI  
ESENDRNKNVFSNSQTMQAL"  
152 AGGGCGAATACCTGCCGGCGAAGAAAGCAAGATCTCATCTTCACACCGGATTCGGTC 211  
11  
3102 ACGGAAATTTGTCCCAAGAACAGCAAGTTCGGTCTATGACACGCGCTCTT 3161  
11  
212 ATTCCGATCTGACCTACCGCTGCGCATATATGCGACGCGCAACATCTTCGGCTCGCGC 271  
11  
3162 ACGGAGACGGTGTGTCGAAGGCAATACACACGACGCTGTCGAATTAAGG 3221  
11  
272 ACCACCTGACCGGTGCTGCGACGGGGCGTCCAAAGTGGCGCTGACCGCGGGTACAGCA 331  
11  
3222 AGCAGCTGACAGGCTTTATGACTCGGCAAAAGCAATGCCAATGACATACCCGTCACAA 3281  
11  
332 AGGACGACTGCGCGATCACCAAGATGCGTGTGATGTGCGACGTGCGCGAATCGT 391  
11  
3282 GAGAGGAATGACCGAAATATCTTGAGAGCCCTCAGAAAAACAACCTCAAGATGCTT 3341  
11  
392 TCGTGAATCTGACCGTCACCGCGGGATACGGAAGCGCAAGGGCGGAAAGACTGTCCA 451  
11  
3342 ACATCGCTCCCACTCAAGGAAATCGGCACTCGGGCTTGACCCCGCTAAATGTGG 3401  
11  
452 AGCTCAACCATGAGGTACATCTAGACCATCCGTAAGTGTGGGCTTCCCGCCGCGC 511  
11  
3402 GGAACCAAGT-----TCATGTTATTTCCTCCAGGGCTGGGGCCATGTACGAG 3452  
11  
512 AGCAGATCTTCGCGACACCGCGATCGCGCGCATCTCGCGCGCGCGCGCGACACA 571  
11  
3453 ACCTTACGAAGTAGGCGTTACCGGAGTTAGCGTGTGTCGGAGAAATGCCCCGGATG 3512  
11  
572 CCGTCGACCGGACATCAAGAACTACCACTGGGGGTATCTCACCCGACGCAATTCGAG 631  
11  
3513 CCCTGTCCCGCAACATCACTCCCTGAACATCTCAACAACATCTCGCAAGATGAGAG 3572  
11  
632 CCAAGGACGCTGTGCGCGACCGCGATCGTGTGCACTCGCAACTGCGAGTGGCCGAG 691  
11  
3573 CCAACGAGGAGGAGAGAGAGATATTTCTTACCATACAGGGGTTGTGTGGAG 3632  
11  
692 GTCCGGGCTTCAACGTGTGATCTCAAGAGCGGCAAG---CTGGCCTCCCGTCCCGGA 748  
11  
3633 GCTCGGGAGACAACATCTGTGTCAAGGACGACAGGGCTCTGACCCCTTCAATATA 3692  
11  
749 ACCGCTTCCGGCGGACACCGCTAAGCGGTGTGCAAGTGGCGGACGATGGGATCG 808  
11  
3693 GCAACCTGAAAGATATACAAAGGCCACACCATCTGCACTTGTGAGCAAAATGGATATA 3752  
11  
809 AAGCCACCTGCGCGACGTCACAGCCGTGACTACAGACGCGGAGAGTATGATGGCG 868  
11  
3753 AGGTATTCGAATCAAACTCGGCGCTGTTGACCTTTACACAGAGATGAGATCTTTGTCA 3812  
11  
869 TCACCAACCGCGGGGGGTACACCGATCAATCGCTGGATGGCGAGCGGTGGGCAAG 928  
11  
3813 CAGAACTGCAAGCGAATCGGCCCGCTGACAGACTTGACGGAAGAGTATAGTACAG 3872  
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

QY 929 GCGAGCCCGGTCACACTGACGCTGCCCAT 956  
 DB 3873 GCAAAACCGCGGCTCTCACAATGAATAAT 3900

RESULT 13  
 PAU63816

LOCUS PAU63816 7991 bp DNA linear BCT 28-AUG-1998

DEFINITION Pseudomonas aeruginosa glnE gene, partial cds; llyE, ADP-heptose:LPS heptosyltransferase I homolog (waaf), lipopolysaccharide heptosyltransferase I homolog (waac), glucosyltransferase I homolog (waag), RfaF protein (waaf), and unknown protein (waax) genes, complete cds, and inaA gene, partial cds.

ACCESSION U63816 GI:3482878

VERSION U63816.1

KEYWORDS Pseudomonas aeruginosa.

SOURCE Pseudomonas aeruginosa.

ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 7991)  
 Franklund, C.V., Coyne, M.J. and Goldberg, J.B.  
 Clustering of the lipopolysaccharide Core Genes, waaf, waac, waag, and waax, in Pseudomonas aeruginosa

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 7991)  
 Coyne, M.J.

AUTHORS Direct Submission

JOURNAL Submitted (15-JUL-1996) Channing Laboratory, Harvard Medical School/Brigham & Women's Hospital, 180 Longwood Avenue, Boston, MA 02115, USA

REFERENCE 3 (bases 1 to 7991)  
 Coyne, M.J.

AUTHORS Direct Submission

JOURNAL Submitted (28-AUG-1998) Channing Laboratory, Harvard Medical School/Brigham & Women's Hospital, 180 Longwood Avenue, Boston, MA 02115, USA

REMARK Sequence update by submitter  
 On Aug 28, 1998 this sequence version replaced gi:2226274.

COMMENT location/Qualifiers

FEATURES

source

1. 7991

/organism="Pseudomonas aeruginosa"

/strain="PAK"

/serotype="O6"

/db\_xref="taxon:287"

/note="Fisher ImmunoType IT1; rfa operon"

<1. 1522

/gene="glnE"

<1. 1522

/gene="glnE"

/note="similar to glnE from a variety of prokaryotic cells"

/codon\_start=2

/transl\_table=11

/product="unknown"

/protein\_id="AAC33171.1"

/db\_xref="GI:3482881"

/translation="1PLMEALDEESASROLADAGFVDAEAMKRLSDLRHGPVRA  
 ORLGRRLDAFVRLAMTVENPDVLEERYLPVEAVARSATVLLTNPGLER  
 LTLCAAPMAVEOIAARPTLLDELINGRLEFRPOAAEIAELERLMRIPEDELE  
 OMETLRHETLAGLTVASSETGTPPLKVSPTLTLWLEALIVELELAKMQLVORH  
 RPLRADGTPCDEVTYVIGKVGLEFHSGLDVFTHDSDPQCEYDGSSTDAOF  
 FTRLOKLIHFLTAQPSGTYLEVMRPSCAGLTVSSLGAFQROEOAWTEHO  
 ALVRAVLACGRVQASFEAVRAEVLARPDALRTVESEKARAKRNLCGRATVAG  
 TASNAFEATAADLKHDAAGIVDIEFVQVYLAWSGEHPALLETDTNIRILEGIERA  
 GLIASEDVRLIOEAKYRAAHLRLALDKAGVYSGEHFOTBRREVITIMWELRG"

1578. 2501

/gene="llyE"

/gene="llyE"

1578. 2501

/gene="llyE"

/note="similar to llyE from Escherichia coli and other

organisms"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAC33172.1"

/db\_xref="GI:3482882"

/translation="MSMADRDGVWYDEGLVOMRDATTHVLTHTLHYGSGVEGVRAY  
 DYPQCTAIFRLQAHTRDLRFLDSAHINNMQIPYSRDINETAAPCARATWKAIPARAC  
 SNGSEMGILRASGLKVHVIYIASMSGAVYGEALQOGIKVRSSTSRHHVYISATRK  
 SNGAYINSLALQEAISGADPAMLDDEGVYVABSSGNITITDGVITYTEVVALN  
 GTRNTITLTAHEHFVKVERKITDEVTYIADAEFTGTAAEVPIPIREVDRKIGAGR  
 RKPVEKLOKAFEDLVSGTEHAHMRTLVK"

2561. 3598

/gene="waaf"

2561. 3598

/gene="waaf"

/note="similar to Escherichia coli RfaF protein encoded by  
 Genbank Accession Number U00039"

/codon\_start=1

/transl\_table=11

/product="ADP-heptose:LPS heptosyltransferase I homolog"

/protein\_id="AAC33166.1"

/db\_xref="GI:3482879"

/translation="WRLIVGSPWGMVMAQTLFQCLKORHPECVIDVLAPEMSRPI  
 LERMEVQALSFPLGHVMDVATRRRIGRIGQYEQAILPLPSLSALVPMFAGIP  
 KRTGRCMRIGLNDIRKLDKORPILIEFMALAFEPGVLEPKPPOPLRIDDS  
 ROMALDKFALSIDRVLALCPGAEGRGKRPAPAHVAANVAEAKIRAGOMVLRGSKD  
 HPGGEIRORLIPGIRRESFNLAGEPTSLAEIIDLMSGAGVYSDSLMVAALDRP  
 LVGYGTSIPQFTPLADRVETVRLGLECSPEFCRTCFGHYNCLRELPGELYQALE  
 RLVGPAFVAG"

3595. 4662

/gene="waac"

3595. 4662

/gene="waac"

/note="similar to Escherichia coli RfaC protein encoded  
 by Genbank Accession Number U00039"

/codon\_start=1

/transl\_table=11

/product="lipopolysaccharide heptosyltransferase I  
 homolog"

/protein\_id="AAC33167.1"

/db\_xref="GI:2226276"

/translation="WRVLVYTSLSGVHTLPALNDARAIRGIDFQVWVEGFAEI  
 PAMPAVAARVIPAIRRMKRLMOTLRNGEWRKQKLAKEDVDLVDAQGLKSAVL  
 TRYVGTIPVAGLSDASAREPLASRTYRAAPVAMGAVERTROLFEQALDVPESV  
 GDYGLDRQLADADPGAPVLYELHGTGTWRDHPAEAVWRLEARMCEGRSGVPLPS  
 AABERARGLAAGLENAAVLPRLSIAMAVYLACARCAVADTGLHARLADVPPLS  
 LKGPNTGFTCAVRSOVYHLSGDPFCAPCLKTCTTYQPTEDRKLFDLKRQPLCFTR  
 LNPQVATQLEAMLLAPETLR"

4659. 5780

/gene="waag"

4659. 5780

/gene="waag"

/note="similar to Escherichia coli RfaC protein encoded by  
 Genbank Accession Number U00039"

/codon\_start=1

/transl\_table=11

/product="glucosyltransferase I homolog"

/protein\_id="AAC33168.1"

/db\_xref="GI:2226277"

/translation="MTLAFLLVYFPFGQLQDFMRLAIECORGHDIRVYTLIMEGD  
 VPDEEVLAAPVSIENHRNKEFTFAVADLDRPQVATIGFNKMGADVVYAADAC  
 FEEFAQTLRNPLRYRHFAGYEAADVDPASKEITLMSVQOPLFKVHGTQAE  
 RFLHLLPGISODRRAPNADVAEFPREFGEEDLLVQIGSGFKTGGLDSKAL  
 ALAPKALRRRTLRILAGODDPKFLIOALAGLNDVOYLIKGSDDIPRLDLADLIL  
 PAYENNGVYLLAIVSGLEFVUTDVCGRATYTAEDAGRVLPSPEDUSLNLRLAEH  
 LEDAPAAASRKLGAADRADLTSMPORADLIDEAS"

5777. 6583

/gene="waap"

5777. 6583

/gene="waap"

/function="putatively involved in phosphorylation of LPS  
 core heptose"

```

/note="similar to Escherichia coli RfaP protein encoded by
GenBank Accession Number U00039"
/codon_start=1
/transl_table=11
/product="RfaP protein"
/protein_id="AAC33169.1"
/db_xref="GI:2226278"
/transl_table="MRLVLEPPKRLMSGDPFEAVEALOGKVFYHELEGRRLRTAEVD
GRGYFKIRHGIGMEIRIAKNLTLAKLPVIGAROEOMAIRLHEAGVTWTAVAGERG
SPAROHSPITVIELAPYDLVEFSODMERPPRPRLKRALVEAVARVMGDMHRAGVN
HRDQYICHELHLDKFPVSDDFRLSTYIDHROTADPATPKRRNDLALYTSALDIDG
LRRDRKRLRFLRYFRRLRLILDEAGLILAMERAKLYEKRRQIGLL"
583. 7317
/gene="waax"
/583. 7317
/gene="waax"
/583. 7317
/note="similar to waap"
/codon_start=1
/transl_table=11
/product="unknown protein"
/protein_id="AAC33170.1"
/db_xref="GI:3482880"
/transl_table="MAGWTLEPPYRHLADFGSDSVFALIEGRRLTRDPLSEYIRIER
DGYRYVYKRYRSAGKLRRLYLRPIKAERONIKREKRGITPAEYVANGLERKAGAP
ORGMATRELPGTEDELVLACNRDPRKDPKRWMDVSKOIAKTRIMHDOHTNDLK
WRMLVDAQPLVYFIDCPNGAFWMSFRLRLRTIKDLACIDKAKYHLSTRTORLEFLYIQ
YRRRLHNESSDKRIRQVAVFEFGRE"
7314. >7991
/gene="inaa"
/7314. >7991
/gene="inaa"
/note="similar to inaa from Escherichia coli protein
encoded by GenBank Accession Number AE000313"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC33173.1"
/db_xref="GI:3482883"
/transl_table="MSVFIAAOHDILORNRLLSPDALMALQIDAVDEPTEKRGWSS
VYRLDIDSAVYLRKOSNLRLSRHLRPLGPTPARERPRERIRALGIPALQAKFAFG
OAVNGEROALITRALDQMDLDTWLSHMDLAEERRRRLRAVGRARGLHQAQGMH
GCTPKHVFLEREGADVFEEKCLDLEKTRITWRGRDQVYDLERMLRRADWSEADVH
EFISAYVD"
BASE COUNT 1330 a 2690 c 2667 g 1304 t
ORIGIN
Query Match 8.2% Score 83.2; DB 1; Length 7991;
Best Local Similarity 48.1% Pred. No. 0.011;
Matches 337; Conservative 0; Mismatches 348; Indels 15; Gaps 3;
QY 256 ATCTTCGGCTCGGGACCACTGACCGGTTGCTCGACGGGGGCTCCAGCTCGGCTG 315
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 1731 ATCTTCGGCTCGGGACCACTGACCGGTTGCTCGACGGGGGCTCCAGCTCGGCTG 1790
QY 316 GACGGCGGCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 375
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 1791 CAGATCCCGTACAGC-----CGGACGAGATCAACGAGGCGCGCGCGCGCG 1841
QY 376 CAGCTGCGGAGTCTGCTGTAATCTGACCGCTACCGGGGATACGAAAGCGCAAGGCG 435
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 1842 GCGAGAACACTCGAAGAACCGGCTATATCCGCCGATGTGTTCAGGGAAGCAAGGCG 1901
QY 436 GAGAAAGACCTGTCCAGCTCAACCATAGGTATCATCTACATCCGCTACTGTGG 495
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 1902 ATGGGCGCTCGGGCGGCGGCTGAAGGTCAATGTATCATCGCTCTGAGCTGGGGC 1961
QY 496 GCGTTCGCCCGCGCGA---GGAGATCTTGGGACACACCGGATCGTGGCGGCATGTC 552
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 1962 GCGTACATGGGAGAGAACCTTCGACGACGACGACGACGACGACGACGACGACG 2021
QY 553 CGCGCGCGCGCGCGCAACCGCTCGACCGGACCATCAAGAACTACAGTGGGATGTC 612
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2022 CGCACACAGTCAACATCTCGATGACCGCGTCCAGTCAAGCGCGCTACATCAATCG 2081

```

```

QY 613 ACCGACGCCAGTTTCGAAACCAAGACCGTGTGCGCGCACCGCATCTCTGCACTCG 672
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2082 ATGCTGGCCCTCCAGAGAACGATCTCCGGCGCGCGACGAGCGCATATGCTCATCCG 2141
QY 673 GACAACTGGTGGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 732
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2142 GAAGGTACGTGGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 2201
QY 733 GCGTCCCGCTCCCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 789
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2202 TACACCGCGGAGTACACCGCTCGCGGACGACGACGACGACGACGACGACGACG 2261
QY 790 GCGGACGAGTGGGACGACGACGACGACGACGACGACGACGACGACGACGACG 849
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2262 GCGGCGGACGAGGTTTAACTGTGTGAGAAAGCGATACCGGACGACGATGATAC 2321
QY 850 GCGGACGAGTGGGACGACGACGACGACGACGACGACGACGACGACGACGACG 909
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2322 GCGGACGAGGCTTCTTCACTGGCACTGCGCGGAGTACACCGCATCCGCAAGTGAC 2381
QY 910 GCGGACGCGGTGGGCAACGCGGACGCGGCGGTCACGACG 949
||||| 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 2382 GGTGCAAGATGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2421
RESULT 14
AE004996 10196 bp DNA linear BCT 12-FEB-2001
LOCUS Halobacterium sp. NRC-1 section 27 of 170 of the complete genome.
DEFINITION AE004996 AB004437
ACCESSION AE004996.1 GI:10580000
VERSION
KEYWORDS
SOURCE
ORGANISM
Halobacterium sp. NRC-1.
Halobacterium sp. NRC-1
Archaea: Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
1 (bases 1 to 10196)
Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sirogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Weir,R., Goo,Y.A.,
Leithausen,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddock,D.G., Jablonski,P.E., Krebs,M.P., Angewine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischroed,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dassarma,S.
Genome sequence of Halobacterium species NRC-1
Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
11016950
2 (bases 1 to 10196)
Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
Shukla,H.D., Lasky,S.R., Balliga,N., Thorsson,V., Sirogna,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Weir,R., Goo,Y.A.,
Leithausen,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddock,D.G., Jablonski,P.E., Krebs,M.P., Angewine,C.M., Dale,H.,
Isenbarger,T.A., Peck,R.F., Pohlischroed,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
Dassarma,S.
Direct Submission
Submitted (14-JUL-2000) Institute for Systems Biology, 4225
Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES
source
1. 10196
/organism="Halobacterium sp. NRC-1"
/strain="NRC-1"
/db_xref="taxon:64091"
257..376
/gene="trpG2"
/note="VN0386"
257..376
CDS

```







```

Db 9656 CTTCTGATGAATAATGGATACAGTTATGAAGCAAAACCTCGCCCTTTTGTGACTCTAC 9715
QY 847 GAGCGGACGAGTTGATGGGGGTCACCAACCGGGGGGTCACACGATCAACTGCTG 906
Db 9716 ACGGACAGACGAAATTTTCGTTACCGGACTGCAGCCGAATCTGCTCGGTCACACGAGCTC 9775
QY 907 GATGGCGAGGCCGTCGAGCAACGCGGAGCCCGGTCACATGACGCTGCCAT 956
Db 9776 GACGGAAGATCATCGGTATGGCAACCGGTCCTCTCAGATGAAGAT 9825

```

Search completed: April 24, 2003, 22:21:04  
 Job time : 3997 secs





GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:00:47 ; Search time 279 Seconds

(without alignments)  
8233.112 Million cell updates/sec

Title: US-10-067-291-2

Perfect score: 1020

Sequence: 1 atgactgcctcttcagacct.....tcgaacgacatcgaatactga 1020

Scoring table: IDENTITY\_NTC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	100.0	1020	21	AAA94032
2	1018.4	99.8	1020	19	AAA94040
3	326.6	32.0	993	19	AAV83675
4	326.6	32.0	993	19	AAV83676
5	326.2	32.0	978	19	AAV73146
6	325	31.9	1574	19	AAV73156
7	66.2	6.5	15872	18	AAV68715
8	66.2	6.5	15872	21	AAV68723
9	64.2	6.3	58857	21	AAA58471

10	59	5.8	1208	15	AAQ64204
11	58.4	5.7	9169	24	ABK91617
12	58.4	5.7	9785	24	ABK91624
13	57.8	5.7	12588	15	AAQ63293
14	57.4	5.6	5392	15	AAQ64201
15	57.2	5.6	2271	11	AAQ68844
16	57	5.6	1140	19	AAV41733
17	56.2	5.5	1053	22	AAH44045
18	56.2	5.5	1065	22	AAH78255
19	56.2	5.5	1065	22	ABLA1306
20	56.2	5.5	6798	22	AAH44043
21	56.2	5.5	6798	22	AAH78258
22	56.2	5.5	6798	24	AAH78258
23	56.2	5.5	8077	24	AAH78258
24	55.2	5.4	2821	22	AAI66075
25	55	5.4	2291	9	AAH80309
26	54.6	5.4	1162	21	AAH89451
27	53.8	5.3	1294	20	AAZ06824
28	53.8	5.3	1294	22	AAH74537
29	53.6	5.3	2634	14	AAQ39093
30	53.4	5.2	12152	22	AAH80699
31	52.8	5.2	1509	24	ABL39958
32	52.4	5.1	810	24	ABK73201
33	51.6	5.1	3412	20	AAK75772
34	51.6	5.1	3756	18	AAH72684
35	51.6	5.1	27541	22	AAH7185
36	51.6	5.1	125401	22	AAH7185
37	51.2	5.0	1509	21	AAH51626
38	50.6	5.0	825	22	AAH00985
39	50.6	5.0	23666	12	AAO10190
40	50.4	4.9	1050	22	AAH41175
41	50.4	4.9	1086	21	AAH89448
42	50.2	4.9	1835	21	AAA30798
43	50.2	4.9	48328	22	AAH28540
44	50	4.9	732	21	AAA39825
45	50	4.9	29879	14	AAQ46806

## ALIGNMENTS

RESULT 1	AAA94032	standard; DNA; 1020 BP.
ID	AAA94032	
XX	AAA94032;	
AC	15-JAN-2001	(first entry)
XX		
DE	Mycobacterium aurum stereoselective transamination catalyst gene.	
XX		
KW	Transamination; optically active amino compound; isomer; ds.	
XX		
OS	Mycobacterium aurum.	
XX		
PH	key	Location/Qualifiers
FT	CDS	1..1020
FT		/*tag= a
FT		/product= "stereoselective transamination catalyst"
XX		
PN	EP1038953-A1.	
PD	27-SEP-2000.	
XX		
PF	14-MAR-2000; 2000EP-0400701.	
XX		
PR	19-MAR-1999; 99UP-0075511.	
PR	30-MAR-1999; 99UP-0088634.	
XX		
PA	(SDMO ) SDMITOMO CHEM CO LTD.	
XX		
PI	Takashima Y, Mitsuda S, Wieser M;	
XX		

snac gene encoding  
Modified HIV prote  
Sequence encoding  
Sequence encoding  
Amylase gene from  
Codon-optimised RA  
Streptomyces sp. C  
Nucleotide sequenc  
Streptomyces IPP r  
Streptomyces sp. C  
Nucleotide sequenc  
Streptomyces sp. C  
Operon F DNA encod  
Triticum aestivum  
Entire amylose gen  
Rice branched chal  
Streptomyces albid  
Nucleotide sequenc  
Streptomyces nodos  
M. carbonacea DNA  
Synthetic Gag poly  
Bacillus lichenifo  
S. erythraea erythr  
Sugar biosynthesis  
Streptomyces nours  
Streptomyces nours  
HIV codon-optimiz  
Mycobacterium aviu  
Cephalosporin ant  
Mammalian Cre reco  
Branched chain am  
DNA encoding maize  
Genomic fragment #  
Synthetic DNA CAH-  
erya region of S.

DR WPI: 2000-629576/61.  
 DR P-PSDB: AAB26100.  
 XX Novel Mycobacterium protein capable of catalyzing transamination stereo  
 PT selectively, useful for producing optically active amino compounds,  
 PT comprises a sequence of 339 amino acids  
 XX  
 PS Claim 7: Page 33-38; 54pp; English.  
 XX The present sequence is the coding sequence for a protein from  
 CC Mycobacterium aurum which acts as a biocatalyst. It is able to catalyse  
 CC transamination stereoselectively to produce an optically active amino  
 CC compound.  
 XX  
 SO Sequence 1020 BP; 198 A; 350 C; 313 G; 159 T; 0 other;

Query Match 100.0%; Score 1020; DB 21; Length 1020;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-186;  
 Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACTGCTCTTTCAGACCTCGGACCTCCAACTGTGGCCGTCGAGCCCGGCGCATC 60  
 DB 1 ATGACTGCTCTTTCAGACCTCGGACCTCCAACTGTGGCCGTCGAGCCCGGCGCATC 60  
 QY 61 CGCGAGGACACCCCGCGCGCTGCTGATCCAGTACAGCAGTACGAACTGGACATCC 120  
 DB 61 CGCGAGGACACCCCGCGCGCTGCTGATCCAGTACAGCAGTACGAACTGGACATCC 120  
 QY 121 AGCCGCTGCGCGGCGCGCTGCTGATCCAGTACAGCAGTACGAACTGGACATCC 180  
 DB 121 AGCCGCTGCGCGGCGCGCTGCTGATCCAGTACAGCAGTACGAACTGGACATCC 180  
 QY 181 AAGATCTTCATCTTCAGACCGGATGCTGCTATTCGATCTGACCTACCGTGCAGAT 240  
 DB 181 AAGATCTTCATCTTCAGACCGGATGCTGCTATTCGATCTGACCTACCGTGCAGAT 240  
 QY 241 GTAGGCGACGGGCAACATCTTCGCGCTGCGGACCACTGAGCCGCTGTCGAGCGGCG 300  
 DB 241 GTAGGCGACGGGCAACATCTTCGCGCTGCGGACCACTGAGCCGCTGTCGAGCGGCG 300  
 QY 301 TCCAAGCTGCGCGCTGAGACCGCGGCTACAGCAAGACGAACTGGCGGATCACCAGAG 360  
 DB 301 TCCAAGCTGCGCGCTGAGACCGCGGCTACAGCAAGACGAACTGGCGGATCACCAGAG 360  
 QY 361 TCGCTGTCGATGTCGACGCTGCGGAACTGCTTCGATCTGACCTGACCGGCGATAC 420  
 DB 361 TCGCTGTCGATGTCGACGCTGCGGAACTGCTTCGATCTGACCTGACCGGCGATAC 420  
 QY 421 GGAAGGCGCAAGGCGGAGAGAGACCTGTCGAACTGACGCTGACGCTGACATCTACGCC 480  
 DB 421 GGAAGGCGCAAGGCGGAGAGAGACCTGTCGAACTGACGCTGACGCTGACATCTACGCC 480  
 QY 481 ATCCCGTACCTGTGGGCTTCCCGCGCGGAGAGATCTTGGCGACACCGCGGATGTC 540  
 DB 481 ATCCCGTACCTGTGGGCTTCCCGCGCGGAGAGATCTTGGCGACACCGCGGATGTC 540  
 QY 541 CCGCGCATGTGCTCCCGCGCGCGGCGCAACGCTGACCCGACATCAAGAACTACAG 600  
 DB 541 CCGCGCATGTGCTCCCGCGCGCGGCGCAACGCTGACCCGACATCAAGAACTACAG 600  
 QY 601 TGGGGTATGTCACCGGAGCGATTTGGAAGCAAGAACCGTGGTGGCGGACCGCGATC 660  
 DB 601 TGGGGTATGTCACCGGAGCGATTTGGAAGCAAGAACCGTGGTGGCGGACCGCGATC 660  
 QY 661 CTGCTGACTGAGACACTGTCGTCGCGGAGAGTCCGGGCTTCAACGTCATGTCAG 720  
 DB 661 CTGCTGACTGAGACACTGTCGTCGCGGAGAGTCCGGGCTTCAACGTCATGTCAG 720  
 QY 721 GACGCAAGCTGCGCTCCCGTCCCGGAGCGGTCGCGGCGGATCAACCCGTAAGAGGTC 780  
 DB 721 GACGCAAGCTGCGCTCCCGTCCCGGAGCGGTCGCGGCGGATCAACCCGTAAGAGGTC 780  
 QY 781 TTGCAACTGGCCGACGACGATGGGATGGAAGCAACCTGCGGACGTCACACCGCTGAA 840

DB 781 TTGCAACTGGCCGACGACGATGGGATGGAAGCAACCTGCGGACGTCACACCGCTGAA 840  
 QY 841 CTCTACAGCGCGGACGAGTGTGAGGGGTACACCGCGGGGGGTACACCGATCAAC 900  
 DB 841 CTCTACAGCGCGGACGAGTGTGAGGGGTACACCGCGGGGGGTACACCGATCAAC 900  
 QY 901 TCGCTGATGCGGAGCGCGGTCGCAACGCGGCGGTCACAGTGCAGTCCG 960  
 DB 901 TCGCTGATGCGGAGCGCGGTCGCAACGCGGCGGTCACAGTGCAGTCCG 960  
 QY 961 GACCGGTTCTGGCGCTGATGACGAGCGGCGCGCTGATGAAAGATATCTGA 1020  
 DB 961 GACCGGTTCTGGCGCTGATGACGAGCGGCGCGCTGATGAAAGATATCTGA 1020

RESULT 2  
 AAA94040  
 ID AAA94040 standard; DNA; 1020 BP.  
 XX  
 AC AAA94040;  
 XX  
 DT 15-JUN-2001 (first entry)  
 XX  
 DE Mycobacterium aurum stereoselective transamination catalyst gene mutant.  
 XX  
 KM Transamination; optically active amino compound; isomer; mutant;  
 KM muten; ds.  
 XX  
 OS Mycobacterium aurum.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1020  
 FT /\*tag= a  
 FT /product= "mutant stereoselective transamination  
 FT catalyst"  
 FT misc\_difference 4  
 FT /\*tag= b  
 FT /note= "wild-type A substituted by G"  
 XX  
 PN EP1038953-A1.  
 XX  
 PD 27-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000EP-0400701.  
 XX  
 PR 19-MAR-1999; 99JP-0075511.  
 PR 30-MAR-1999; 99JP-0088634.  
 XX  
 PA (SUMO) SUMITOMO CHEM CO LTD.  
 XX  
 PI Takashima Y, Mitsuda S, Wleser M;  
 XX  
 DR WPI: 2000-629576/61.  
 DR P-PSDB: AAB26104.  
 XX  
 PS Claim 8: Page -, 54pp; English.  
 XX  
 CC The present sequence is a mutant version of the coding sequence for a  
 CC protein from Mycobacterium aurum which acts as a biocatalyst. It is able  
 CC to catalyse transamination stereoselectively to produce an optically  
 CC active amino compound.  
 CC Note: This sequence is not shown in the specification, but is derived  
 CC from the Mycobacterium aurum wild-type sequence shown in SEQ ID NO: 2  
 CC (see AAA94032)1.  
 XX  
 SO Sequence 1020 BP; 197 A; 350 C; 314 G; 159 T; 0 other;

Query Match 99.8%; Score 1018.4; DB 21; Length 1020;

Best Local Similarity 99.98; Pred. No. 5e-186;  
Matches 1039; Conservative 0; Mismatches 1; Indels 0; gaps 0.

OY	1	ATGACTCTCTTTAGACCTTCGAGACCTTCACACTGGTGGCCGTGACGCCGGCGCATC	60
OY	1	ATGCGTCTCTTTAGACCTTCGAGACCTTCACACTGGTGGCCGTGACGCCGGCGCATC	60
Db	1	ATGCGTCTCTTTAGACCTTCGAGACCTTCACACTGGTGGCCGTGACGCCGGCGCATC	60
OY	61	CGCGAGACACCCCGCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACTCC	120
Db	61	CGCGAGACACCCCGCGGCTCGGTGATCCAGTACAGCGACTACGAACTGGACACTCC	120
OY	121	AGGCCGTTCCGCGCGCGCGTCCGCTGGATCGAGGGCGCAATCTGCGCGGCCGAAGGCG	180
Db	121	AGGCCGTTCCGCGCGCGCGTCCGCTGGATCGAGGGCGCAATCTGCGCGGCCGAAGGCG	180
OY	181	AAAGTCTCCATCTTCGACACCCGATTCGGTATTCGATCTGACCTTACACCGTCCGCAT	240
Db	181	AAAGTCTCCATCTTCGACACCCGATTCGGTATTCGATCTGACCTTACACCGTCCGCAT	240
OY	241	GTAAGCGACGGCAACATCTTCGCGGCTGGGGGACCACTGGACCGGGTTGCTCGAGGGGGG	300
Db	241	GTAAGCGACGGCAACATCTTCGCGGCTGGGGGACCACTGGACCGGGTTGCTCGAGGGGGG	300
OY	301	TCCAAAGTCGCGCTTGAGACGCCGGGTACAGCAAGAGACAACTGGCCGAGATCAACAAG	360
Db	301	TCCAAAGTCGCGCTTGAGACGCCGGGTACAGCAAGAGACAACTGGCCGAGATCAACAAG	360
OY	361	TGCGTGTTCGATGTGCGACGTGCGGAAATCGTTGCGAATCTGACCGTTCACCCGGGGATAC	420
Db	361	TGCGTGTTCGATGTGCGACGTGCGGAAATCGTTGCGAATCTGACCGTTCACCCGGGGATAC	420
OY	421	GGAAGCGCAAGGGCGAGAAAGAACTTCACAGCTCAACCATAGATGATCATCTACGCC	480
Db	421	GGAAGCGCAAGGGCGAGAAAGAACTTCACAGCTCAACCATAGATGATCATCTACGCC	480
OY	481	ATCCCGTACCTGTGGGCTTCGCCGCCCGCGAGAGATCTTCGGGACACACCGGATCGTG	540
Db	481	ATCCCGTACCTGTGGGCTTCGCCGCCCGCGAGAGATCTTCGGGACACACCGGATCGTG	540
OY	541	CCGCGCACTGTCCGCGCGCGCGCGCGCAACCGTTCACACCGCACTCAAGACTACAG	600
Db	541	CCGCGCACTGTCCGCGCGCGCGCGCGCAACCGTTCACACCGCACTCAAGACTACAG	600
OY	601	TGGGGTGTCTCACCGCACCGCAATTTGAAAGCCAAAGACCGTGGTGGCGCACCGGATC	660
Db	601	TGGGGTGTCTCACCGCACCGCAATTTGAAAGCCAAAGACCGTGGTGGCGCACCGGATC	660
OY	661	CTGCGCGACTCGGACCACTGCGTGGCGGAAGTCCGGGCTTCAACGTTGATCTGATCAAG	720
Db	661	CTGCGCGACTCGGACCACTGCGTGGCGGAAGTCCGGGCTTCAACGTTGATCTGATCAAG	720
OY	721	GACGGCAAGCTGGGCTCCCGCTCCCGGACGCGTTGGCGGGCATCACCCGTAAAGCGGTG	780
Db	721	GACGGCAAGCTGGGCTCCCGCTCCCGGACGCGTTGGCGGGCATCACCCGTAAAGCGGTG	780
OY	781	TTGCAACTGGCGACGACATGGGATGGAAGCAACCTTGCGGCGACGTACACCAAGCGTGA	840
Db	781	TTGCAACTGGCGACGACATGGGATGGAAGCAACCTTGCGGCGACGTACACCAAGCGTGA	840
OY	841	CTTACGACGCGCGACGAGTTGATGGCGGTACACACCGCGGGCGGGGTACACCGATCAAC	900
Db	841	CTTACGACGCGCGACGAGTTGATGGCGGTACACACCGCGGGCGGGGTACACCGATCAAC	900
OY	901	TGCGCGAATGGCGAGGCGGTGGGGAAGGCGGAGCGCGGCTCACGACGTTGGCATTCGGG	960
Db	901	TGCGCGAATGGCGAGGCGGTGGGGAAGGCGGAGCGCGGCTCACGACGTTGGCATTCGGG	960
OY	961	GACCGGTTCTGGGCGGTGATGAGACGCGGGGCCGCTGATCGAAACGATCGAATCTGA	1020
Db	961	GACCGGTTCTGGGCGGTGATGAGACGCGGGGCCGCTGATCGAAACGATCGAATCTGA	1020

AAV83675  
ID AAV83675 standard; DNA; 993 BP

AC AAV83675

DT 26-FEB-1999 (first entry)

DE DNA encoding a protein with stereoselective transaminase activity

KW Stereoselective; transaminase activity; optically active amine;

KW agrochemical; ds.

05 *Arthrobacter* sp.

Key	Location/Qualifiers
FH	
224	

11  
12  
13

AA  
PN  
W09848030-A1

XX  
PD 29-OCT-1998.

XX 20-APR-1998: 98WO-JP01814

XX 23-APR-1997. 97JP-0121732  
 PR

XX  
XX  
DA / KANE \ KANEKA COPB

XX	Tkachenko V	Tusakov A	Kizak N	Matsu moto K.
DT	Hagocanova T			

PI Ogura M, Yamada Y;  
xy

DR WPI; 1998-583664/49  
DR D-DCDP; 1998-583664/49

XX  
E

PT donor - using a transaminase derived from an Arthrobacter species.

PS Claim 23; page 57; 84pp; Japanese.

CC The present sequence encodes a pro-

CC the optically active amines (having R-chirality) of the

presence of a primary amine as amino donor to produce the

use as pharmaceuticals and agrochemicals.

Sequence 993 BP; 185 A; 324 C; 297 G; 187 T; 0 other;

Query Match	32.08:	Score	326.6:	DB	19:	Length	9
-------------	--------	-------	--------	----	-----	--------	---

```

Best Local Similarity 36.36; Prev. NO: 9.7e-34;
Matches 572: Conservative 0: Mismatches 409: Indels

```

40 GCGGTGAGCCGCGCCATCCGCGAGGACCCCGGCGCGCTCGGTATCCAG

13 GCGGATACCTCCGAGATCTGCTACACGCGACGACCGGCTTCGACTTAA

Query Match	32.08;	Score 326.6;	DB 19;	Length 993;
-------------	--------	--------------	--------	-------------

Matches 572; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY	40	GCCGTCAGAGCCCGGCGCCATCCGGAGAGACACCCCGCGGCTCGGATATCCAGTACAGC	99
Db	13	GGCGATACCTCCCGAGANTGTCTTACACGGACGACACCGGCTCGACTACATCTACTTATAGC	72
QY	100	GACTACGAACCTGGACACCTCCAGCCGTTGCGCGCGGCGGCTCGGCTTGATCGAGCGGAA	159
Db	73	GACTACGAACCTCGATCTCTGCTTACCCCGCTCCGGAGAGTGGCGCATGATCGAGGGTGA	132
QY	160	TACCGCGCGCGCGAGAAAGGACATCTCCATCTTCGACACCGGATTCGGTCAATCCGAT	219
Db	133	TTGCGCCGCGGTGCGAGAGCCCGGATCTTCGATCTTCGATCAGGGTTACCTTCACCTCGGAC	192
QY	220	CTGACCTACACCGTGGCGCATGTATGGCAGGCAACATCTTTCGCGGCTCGGCGACCACTG	279
Db	193	GTCACCTACAGGTCCTTCCAGTCTGCAACGGAATCATCTCCGCTCGACGACACCATTC	252
QY	280	GACCGGTCGTCGAGACGGGCGGCTCAACGTCGGCGCTGGACGCGCGGGTACAGAACAGACAA	339
Db	253	GAGCGCTCTTCTCCACGCGGAGTCAATGCGCATCATCCTCCGCTACACAGGAGAGAA	312

QY 340 CTGGCCGAGATCACCAGAAAGTGCCTGTCATGTCGACAGCTGCGCAATGCTTGATGAT 399  
 DB 313 GTGAAGAGAGATGGCGTGAAGCTGTGGCGAAGACCGAATTGGCTGAGGCTTGCTGTC 372  
 QY 400 CTGACCGCTACCCGGGATACGAAAGCGCAGAGCGCAAGACCTGCTCAAGCTCAC 459  
 DB 373 GTGTGATTAACCGCGGTATAGCTCACTCCGCGGCGAGCGCCATCAGAACACCGC 432  
 QY 460 CATCAGGTGATCTATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGAT 519  
 DB 433 CCGAGAGGTGATCTATGATCCGATCCGATCCGATCCGATCCGATCCGATCCGAT 492  
 QY 520 TTGCGACACACCGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579  
 DB 493 GACGCGGTGACGCGCGATGTCGACAGAGCGGTGCGCGAAGCCCGCGCGCGCGCGCG 552  
 QY 580 CCGACCATCAAGAACTACCACTGAGGAGTCTACCGCGCGCGCGCGCGCGCGCGCG 639  
 DB 553 CCTCAGGTCAAGAACTTCCACTGAGGAGGATCTGATCCGCTTCAAGAGACGCGAC 612  
 QY 640 CGTGGTGGCG 699  
 DB 613 CCGCGGTTCGAGGCTCCCGCTTCTCTCTGACGCGCGCGCGCGCGCGCGCGCGCG 672  
 QY 700 TTCAACGTGTGATCTGTCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759  
 DB 673 TTCAACGTGTGATCTGTCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 722  
 QY 760 GGCATCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCGCG 819  
 DB 733 GGCATCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCGCG 792  
 QY 820 CGCGACGTCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCG 879  
 DB 793 GCGCATCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCGCG 852  
 QY 880 GCGCGGTTCACACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCG 939  
 DB 853 GCGCGGTTCACACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCG 912  
 QY 940 CCACGTCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCG 999  
 DB 913 CCACGTCAACCGCTAGAGCGGTGTCGAAGCTGCGCGCGCGCGCGCGCGCGCG 972  
 QY 1000 ATCGAAGCATCGAATCTGA 1020  
 DB 973 CTTACGCTGTGCGAGTCTGA 993

RESULT 4  
 AAV83676  
 ID AAV83676 standard; DNA: 993 BP.

AC AAV83676;  
 DT 26-FEB-1999 (first entry)  
 DE DNA encoding a protein with stereoselective transaminase activity.  
 XX  
 KW Stereoselective; transaminase activity; optically active amine;  
 KM R-chirality; ketone; synthetic intermediate; pharmaceutical;  
 KW agrochemical; ds.  
 XX  
 OS Arthrobacter sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..978  
 XX /tag= a  
 PN MO9848030-AL.  
 XX  
 PD 29-OCT-1998.

XX 20-APR-1998; 98WO-JP01814.  
 XX 23-APR-1997; 97JP-0121732.  
 PR (KANF) KANEKA CORP.  
 PA Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;  
 PI Ogura M, Yamada Y;  
 PI WPI: 1998-583664/49.  
 DR P-PDB; AAW87564.  
 XX  
 PT Production of optically active amine(s) from ketone(s) and an amino  
 donor - using a transaminase derived from an Arthrobacter species.  
 PS Claim 23; Page 57; 84pp; Japanese.

Query Match 32.0%; Score 326.6; DB 19; Length 993;  
 Best Local Similarity 58.3%; Pred. No. 9.7e-54;  
 Matches 572; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY 40 GCGGTGACCG 99  
 DB 13 GCGGATACCTCGAAGATGCTTCAACGACGACGACCGCGCGCGCGCGCGCGCG 72  
 QY 100 GACTACGAAGTGAACACTTCAAGCCCGTTCGCGCGCGCGCGCGCGCGCGCG 159  
 DB 73 GACTACGAAGTGAAGTCTCTAACCCTCGCGGAGGCGCGCGCGCGCGCGCGCG 132  
 QY 160 TACCTGCG 219  
 DB 133 TTGCTGCG 192  
 QY 220 CTGACCTACACCG 279  
 DB 193 GTCACCTACAGGTCTTCCAGCTGCGAAGCAAGAAATGATTCGCGTGCAGCACATC 252  
 QY 280 GACCGGTCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 339  
 DB 253 GACCGGTCTTCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312  
 QY 340 CTGGCGGATGATCAAGAAAGTGCCTGTCGATGTCGAGCTGCGCGCGCGCGCG 399  
 DB 313 GTGAAGAGATGCGCTCCAGCTGCGCGAAGACCGAATGCGAGCGCTTGTGCG 372  
 QY 400 CTGACCGTACCCGGGATACGAAAGCGCAAGCGCAAGGACCTGCTCCAGCTCAC 459  
 DB 373 GTGTGATTAACCGCGGTATAGCTCACTCCGCGGCGAGCGCGCGCGCGCGCG 432  
 QY 460 CATCAGGTGATCTATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 519  
 DB 433 CCGAGAGGTGATCTATGCGCGTCCCATATGATGATGATGATGATGATGATG 492  
 QY 520 TTGCGACACACCGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579  
 DB 493 GACGCGGTGACGCGCGATGTCGACAGAGCGGTGCGCGAAGCCCGCGCGCGCG 552  
 QY 580 CCGACCATCAAGAACTACCACTGAGGAGTCTACCGCGCGCGCGCGCGCGCGCG 639  
 DB 553 CCTCAGGTCAAGAACTTCCACTGAGGAGGATCTGATCCGCTTCAAGAGACGCGAC 612  
 QY 640 CGTGGTGGCG 699

```

Db 613 CCGGGGTTGAGGCTCCCTTCTGCTCGACGCGAGTGAAGTCTCTGCGGAGGGGCTCGGG 672
Qy 700 TTCAAGCTGTGATCTCAAGAGCGCAAGTGGCCCTCCCGTCCCGAAGCGCTTGGCG 759
Db 673 TTCAAGCTGTGATCTCAAGAGCGCGCTGTCGCGAGCCCGGGTCCGAGCGCGCTCCCG 732
Qy 760 GGCATCACCCTGTAAGAGGCTTCCAGACGCGGCGAGCCAGATGGGATGGAGAGCCAGCTG 819
Db 733 GGCATCACCCTGTAAGAGGCTTCCAGACGCGGCGAGCCAGATGGGATGGAGAGCCAGCTG 792
Qy 820 CCGAGCTGACCAAGCGCTGCTACGACGCGGAGAGTGGATGGAGCGGCTGACCAACCGG 879
Db 793 GCGGACATCAAGCTGCTGTAAGAGCGCGAGCGGAGAGTGGATGGAGAGCGGCTGACCA 852
Qy 880 GCGGAGCTGACCAAGCGCTGCTGTAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 939
Db 853 GCGGAGCTGACCAAGCGCTGCTGTAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 912
Qy 940 CCACTGACGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 999
Db 913 CCACTGACGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 972
Qy 1000 ATCGAAGAGATCGAATCTGTA 1020
Db 973 CTTCAGGCTGTGCGAGTACTGA 993

```

## RESULT 5

AAV73146  
AAV73146 standard; DNA: 978 BP.

AAV73146;

26-FEB-1999 (first entry)

DNA encoding a protein with stereoselective transaminase activity.

Stereoselective; transaminase activity; optically active amine;

R-chirality; ketone; synthetic intermediate; pharmaceutical;

agrochemical; ds.

Arthrobacter sp.

Key Location/Qualifiers  
CDS 1..978  
/\*tag= a

W09848030-A1.

29-OCT-1998.

20-APR-1998; 98MO-JP01814.

23-APR-1997; 97JP-0121732.

(KANF) KANEKA CORP.

Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;

Ogura M, Yamada Y;

WP1: 1998-583664/49.

P-PSDB: AAM82241.

Production of optically active amine(s) from ketone(s) and an amino donor - using a transaminase derived from an Arthrobacter species.

Claim 23; Page 57; 84pp; Japanese.

The present sequence encodes a protein with stereoselective transaminase activity. The transaminase is used to produce the optically active amines (having R-chirality) of the invention. A ketone is treated with a transaminase in the

CC presence of a primary amine as amino donor to produce the  
CC amines. The chiral amines are synthetic intermediates for  
CC use as pharmaceuticals and agrochemicals.

Sequence 978 BP; 183 A; 319 C; 292 G; 184 T; 0 other;

Query Match 32.0%; Score 326.2; DB 19; Length 978;

Best Local Similarity 58.5%; Pred. No. 1.2e-53;

Matches 568; Conservative 0; Mismatches 403; Indels 0; Gaps 0;

```

Qy 50 CCGGGGTTGAGGCTCCCTTCTGCTCGACGCGAGTGAAGTCTCTGCGGAGGGGCTCGGG 109
Db 8 CCGAGATGCTGTAAGAGCGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 67
Qy 110 TTGACACCTTCCAGCGGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 169
Db 68 TTGATCTGCTGTAAGAGCGCGAGCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 127
Qy 170 CCGAAGAGCGAAGATCTGATCTGACACGCGGATGCGGATGCGGATGCGGATGCGGAT 229
Db 128 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 187
Qy 230 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 289
Db 188 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 247
Qy 290 TTGACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 349
Db 248 TTGACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 307
Qy 350 TTGACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 409
Db 308 TTGACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 367
Qy 410 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 469
Db 368 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 427
Qy 470 ACATCTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 529
Db 428 ACATCTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 487
Qy 530 CCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 589
Db 488 ACAGGATGCTGTAAGAGCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 547
Qy 590 AGAATCTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 649
Db 548 AGAATCTACGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 607
Qy 650 GCACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 709
Db 608 AGGCTCCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 667
Qy 710 GCATGCTCAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 769
Db 668 TCGATGCTCAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 727
Qy 770 GTTACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 829
Db 728 GGAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 787
Qy 830 CCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 889
Db 788 CCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 847
Qy 890 CAGGATGCTGTAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 949
Db 848 GCGGATGCTGTAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 907
Qy 950 TTGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1009
Db 908 ACTGATGCTGTAAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 967

```

OY 1010 TCGAATCTGA 1020  
 Db 968 TCGAGTACTGA 978

## RESULT 6

AAV73156  
 ID AAV73156 standard; DNA; 1574 BP.

AAV73156;

26-FEB-1999 (first entry)

PstI fragment of a stereoselective transaminase encoding genomic DNA.

De Stereoselective; transaminase activity; optically active amine;  
 R-chirality; ketone; synthetic intermediate; pharmaceutical;  
 agrochemical, ss.

OS Arthrobacter sp.

PN WO9848030-A1.

PD 29-OCT-1998.

PF 20-APR-1998; 98WO-JP01814.

PR 23-APR-1997; 97JP-0121732.

PA (KANF) KANEKA CORP.

PI Hasegawa J, Ikenaka Y, Iwasaki A, Kizaki N, Matsumoto K;  
 Ogura M, Yamada Y;

DR WPI; 1998-583664/49.

PT Production of optically active amine(s) from ketone(s) and an amino  
 donor - using a transaminase derived from an Arthrobacter species.

PS Example 21; Page 63; 84pp; Japanese.

CC The present sequence represents a 1.6 kb pstI fragment of  
 CC genomic DNA encoding a protein with stereoselective  
 CC transaminase activity. The transaminase is used to produce  
 CC the optically active amines (having R-chirality) of the  
 CC invention. A ketone is treated with a transaminase in the  
 CC presence of a primary amine as amino donor to produce the  
 CC amines. The chiral amines are synthetic intermediates for  
 CC use as pharmaceuticals and agrochemicals.

SQ Sequence 1574 BP; 285 A; 530 C; 476 G; 283 T; 0 other;

Query Match 31.9%; Score 325; DB 19; Length 1574;

Best Local Similarity 58.2%; Pred. No. 2e-53;

Matches 571; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

OY 40 GCGGTGAGCGCGCGCATCCGCGAGACACCGCGCGCGGTGATCCAGTACAC 99  
 Db 398 GCGGATACCTCCGAGATGCTTACACGACGACCGCGCTGACTACTACTATATAC 457  
 OY 100 GACTAGCAACTGACACCTCCAGCCGCTTCGCGCGCGCGGTGATCGAGGCGCA 159  
 Db 458 GACTAGCAACTGATCTCTGATACCGCGCTCGCGGAGGTGCGCATGATGAGGTTGA 517  
 OY 160 TACCTCCCGCGGAGAGACGAGATCTTCATCTTCGACACCGGATTCGGTATCCGAT 219  
 Db 518 TTTCTGTCGCGCGCGGAGGCGGATCTTCATCTTCGATCAGGTTACTTCCATCGGAC 577  
 OY 220 CTGACCTACACCGTGGCGATGATGGACGACGACATCTTCGGGTGCGGACACACCTG 279  
 Db 578 GTACCTACACGCTTTCACGCTTGGAGAGGAAATGCAATTCGCTGACGACACCATC 637

OY 280 GACCGGTGCTCGACGCGGCGTCCAGCTGCCGCTGGACCGCGGGTACAGAACGCA 339  
 Db 638 GAACGCTCTTTCACACCGGAGTGCATGCGCATATCCCTCCGCTCACAGACGCA 697  
 OY 340 CTGCGGAGATCACAAGAGTGTGATGTCGACCTGCGGATCGTTGCTGAT 399  
 Db 698 GTGAGGAGATTCGCTCAACCTGTCGGAAGACGAAATTCGTAAGGCTTCGTCGC 757  
 OY 400 CTGACCGTCAACCGGGGATACGAAAGCCGAGGCGAGAGACCTGTCCAGTCAAC 459  
 Db 758 GTTCGATTAACCGCGGTTCACGCTGACTCGGCGAGCGGACATCCAGAACCGCG 817  
 OY 460 CATCAGGTGATACCTACGACCTCCGATCTGCGGCTTCGCGCGCGCGGACGATC 519  
 Db 818 CCGAGGTGATATGATGATCCGTCATGATGATGATGATGATGATGATGATGATG 877  
 OY 520 TTGCGGACACCGGATTCGCGCGGATTCGCGCGCGCGCGCGCGGACGATGAC 579  
 Db 878 GACGCGGTGACCGGATGCTGCGACAGACGCGCGCGCGCGCGCGCGCGCGCGCG 937  
 OY 580 CCGACCATCAAGAACTACCAAGTGGGATCTCACCGGACGCGGATTTGAGCCAGAC 639  
 Db 938 CCTCAGGTCAAGAACTTCCAGTGGGAGATCTGATCCGCTGCAAGAGACGACGAC 997  
 OY 640 CGTGTGCGCGGACCGGATCTGCTGACCTGCGACAACTGCGTGGCGGACGATCGG 699  
 Db 998 CGCGGTTGAGAGCTCCCTTCGCTGACGCGGATGATGATGATGATGATGATGAT 1057  
 OY 700 TTCAACGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 759  
 Db 1058 TTCAACGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1117  
 OY 760 GGCATCACCCGTAAAGCGGTGTTGAACTGGCGGACGACGATGGGATGAGAGCA 819  
 Db 1118 GGCATTAAGCGGAAAGACGCTGCTGCAAGATCGCCGATGCTGCGACGAGCGAT 1177  
 OY 820 CCGGACGTCACACCGCGGATCTAGACGCGGACGAGTGTGATGATGATGATGAT 879  
 Db 1178 GCGGACATCACGCTGCTGAAAGTGTGCGACCGCGGAGTGTGCGGCTGCAAGC 1237  
 OY 880 GCGGAGTGCACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 939  
 Db 1238 GCGGAGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1297  
 OY 940 CCATGACGTTGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 999  
 Db 1298 CCATGACGTTGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1357  
 OY 1000 ATCGAAAGATCGAATCTGA 1020  
 Db 1358 CTTCAGCTGTGCACTGA 1378

## RESULT 7

AAAT68715  
 ID AAAT68715 standard; DNA; 15872 BP.

AAAT68715;

01-SEP-1997 (first entry)

Streptomyces venezuelae polyketide synthase vep ORF1.

Polyketide synthase; polyhydroxyalkanoate monomer synthase;

polyhydroxybutyrate; biodegradable polymer; vep gene;

metabolic engineering; ss.

Streptomyces venezuelae.

Key Location/Qualifiers  
 CDS 20..13912  
 FT /\*tag= a  
 FT 14056..14136



Query Match	Best Local Similarity	Matches	Score	DB	Length
Query Match	Best Local Similarity	Matches	Score	DB	Length
6.5%;	44.4%;	358;	66.2;	15872;	
		Conservative	Pred. No.	Mismatches	Indels
		0;	0.00063;	443;	6;
				Gaps	
50	CGCGGCCATCCGCGAGACACCCCGCGGCTCGGTGATCGATACGACGACTACGAAC	109			
765	CGCGGCCCTCCGCGAGACGACCGTGTCCACGCGGTCTATCCGCGACGCGCTTCACCA	824			
110	TGGAACTCCGACCCCGCTTCGCGCGCGCGCTGCTGATCGAGGCGCAATACCTCGCG	169			
835	ACGAGGAGACCCCGCGGCTCTACCGTCCCGACGAGGCGCGCCCGAGAGAGTCTTC	884			
170	CGGAG---AAGCGAAGATCTCCATCTTGACACCGAGATTGCTATTCGATCTGACCT	226			
885	GCGAGGCGGACCGGAGGCGCGCTTGACACCGCTCGCGCTCCAGTACGTGCAATCTCAG	944			
227	ACACGCTGCGCATGTATGGACGGCA--ACATCTTCGCGCTCGCGACCACTGGACC	283			
945	GCACCGGAACCCCGCGGCGGACCCCATGAGGCGCGCGCGCGCGCGCTCGCGCT	1004			
284	GGTGTCTGACGCGGCGCTCCAAAGCTTCGCTTGAGCGCGCGGATACAGCAAGCAACTG	343			
1005	CGCGCGCGCGCGCGGACGCAACCCCTGCTGTGCGGCTCGCGCAAGAACGTCGGCGAC	1064			
344	CGGAGTACCAAGAAGTCTGTCGATGTCCAGCTGCGGCAATCTGTCGATCTGA	403			
1065	TGGAGGCGCGCGCGGATACGGAAGCGCAAGCGGAGAGAACTGTTCACACTCACCATC	463			
1125	GGATCCCGCGGAGGCTCACTTCCTCGTACGCCCGACCGGAGATCCGCTGACACCTCG	1184			
464	AGGTGATCTACGCGATCCGCTGATCTGTGGGCTTCCCGCGCGGAGCAAGATCTTG	523			

Db	1185	GGCTCGACGCGCCCGACGGGCTGCGGGAGTGGCGGACCCGGACCGGAACTCTCTCGCG	1244
Oy	524	GCACACCGCGATCTGCGCGCCCATGTCGCGCGCGCGCGCGCAACACCTGACCCGA	583
Db	1245	GGCGTCACTGCTTCGGCATGGCGGCGGACCAACGCCCACTGTGCTTACGGAAGGCCG	1304
Oy	584	CCATCAAGAACTCAAGTGGGGGATCTCACCGCAGCCAGTTTCCGAAGCCAGACCGTG	643
Db	1305	CCGACGGCGGCGGACGACCGCCCGGATCATGAGAGAACCCCGTGCACAGCGGGCGCAC	1364
Oy	644	GTGCGCGCACCGGATCTGCTCTGACTCGAGAACCTGGCTGGCCGAAGTCCGGCTTCA	703
Db	1365	TGCCCTTCGTCTGTCACCGGCGCGGGGGGAGAGCCCTCGCGGCCACGCGCGCGCTG	1424
Oy	704	ACGTGTCATCTGTCGAAGGACGGCAGACTGGCCTCCCGCTCCCGGACGCGTGCAGCA	763
Db	1425	ACGAGGCGCGTCGAAGGGGACCGGAGCTCGCGCCCGGCACTCGCCCGTGGCTGCTCA	1484
Oy	764	TCAACCGTAAGAGCGGTCTGAACTGAGCGGCGACAGATGGGATGAGAGCCAGCCTGGCG	823
Db	1485	CCACCCGTACGCGTCTTCACGCAACCGGCTGCTGCTCGCCCGGACGCGCGCCTCC	1544
Oy	824	ACGTACACACCGCTGAACCTCTACGACG	850
Db	1545	TCGACGGCCTCGGCGGCCCTCGCCCGCG	1571
RESULT	8		
AAZ87283			
ID	AAZ87283	standard; DNA; 15872 BP.	
AC	AAZ87283:		
XX			
DT	05-JUN-2000	(first entry)	
XX			
DE	S. venezuelae vep ORF 1, seq ID NO:1.		
KW	Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;		
KW	neomethymycin; nardomycin; polyhydroxylkanooate monomer synthase;		
KW	biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;		
KW	chronic obstructive pulmonary disease; respiratory inflammation;		
KW	hypercholesterolaemia; crop protection agent; ds.		
XX			
OS	Streptomyces venezuelae ATCC15439.		
XX			
XX	Key	Location/Qualifiers	
FF	CDS	20..13912	
FF		/tag="a	
FF	CDS	/product="vep ORF 1 amino acid sequence #1 (AAAT77177)"	
FF		14056..14151	
FF		/tag="b	
FF	CDS	/product="vep ORF 1 amino acid sequence #3 (AAAT77199)"	
FF		14167..15827	
FF		/tag="c	
FF		/product="vep ORF 1 amino acid sequence #2 (AAAT77178)"	
XX			
XX	MO20000620-A2.		
XX			
PD	06-JAN-2000.		
XX			
XX	25-JUN-1999;	99MO-US14398.	
XX			
XX	26-JUN-1998;	98US-0105537.	
XX			
PA	(MENU ) UNIV MINNESOTA.		
XX			
PI	Sherman DH, Liu H, Xue Y, Zhao L;		
XX	WPI; 2000-160679/14.		
DR	P-PSDB; AAT77177, AAT77178, AAT77199.		
XX			
PT	Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.		







Db 13638 CGACCCGCGGAGGCCCCCTACCCGCCCTGCACCGGCTGTGAGAGTCCGCCGC 13697  
 QY 820 CGCGACGTACACCGCTGAACTCTACGACGCC 852  
 Db 13698 CGCGGCGCGGACGCCCTGCGGTGCTGCGCGGC 13730

RESULT 10  
 AA064204  
 ID AA064204 standard; cDNA; 1208 BP.  
 AC AA064204;  
 XX  
 XX

18-NOV-1994 (first entry)

snac gene encoding enzyme in streptogramin biosynthetic pathway.

XX Antibiotic: streptogramin; snac; snab; snac; biosynthesis; enzyme;  
 KW biosynthetic pathway; Streptomyces pristinaespiralis; ds.  
 XX

OS Streptomyces pristinaespiralis.

XX Key Location/Qualifiers  
 FT CDS 1..1209  
 FT /\*tag- a

FR2696189-A.

PD 01-APR-1994.

25-SEP-1992; 92FR-0011441.

25-SEP-1992; 92FR-0011441.

PA (RHON) RHONE POULENC ROGER SA.

PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P;  
 PI Thibaut D, Zagorec M;

DR WPI; 1994-128286/16.

DR P-PSDB; AAR54204.

XX DNA involved in streptogramin antibiotic biosynthesis - for  
 PT prodn. or bio-conversion of streptogramin(s) or prodn. of  
 PT streptogramin intermediates, derivs. or hybrid antibiotics  
 XX  
 PS Claim 2; Page 54-55; 83pp; French.

XX The snac gene product is involved in the biosynthesis of  
 CC streptogramins, antibiotics active against Gram-positive bacteria.  
 CC The identification of the sequences encoding the enzymes involved  
 CC in the biosynthetic pathway means that they can be isolated and  
 CC manipulated. Mutant microorganisms in which a step in the  
 CC streptogramin biosynthetic pathway is blocked can be cultured to  
 CC produce streptogramin intermediates, which may later be converted  
 CC to streptogramin derivatives. Recombinant cells may also be used  
 CC for the bioconversion of streptogramins from one form to another or  
 CC for the production of hybrid antibiotics.  
 XX

Sequence 1208 BP; 190 A; 500 C; 360 G; 158 T; 0 other;

Query Match 5.8%; Score 59; DB 15; Length 1208;  
 Best Local Similarity 42.3%; Pred. No. 0.014;  
 Matches 390; Conservative 0; Mismatches 530; Indels 3; Gaps 1;

QY 16 GACCTGGGACCTCAACCTGTGTGCGGAGCCGCGGCGCATCCGAGACACCCCG 75  
 Db 73 GACACGCTCTCGACGCGCTGTGCGGAGAGACCCGCTCAGCGCTCGCGGAGAC 132  
 QY 76 GCGGCGCTGTGATCACTACGACGATCAAGAACTGACACCTCCGCTTCCGCGGC 135  
 Db 133 CTGATCACACCGCGGAGGTTCATGCGCGGAGGTGATCAACCAAGCGTACGCGGCC 192

QY 136 GCGGTCCCTGATCGAGGGCGAATACCTGCGGCGGCGAAGACGAAATCTCATCTTC 195  
 Db 193 ATGCCCAACTGTGTCCGCGACAGATCTGGCCATGGGTACGATCGTCGCCCAAGGCG 252  
 QY 196 GACACCGGATTTGGGTATTTCCGATCTGACACTTACACCGTGGCGATGTATGCGGCAAC 255  
 Db 253 TTGAGAGGCGCTCTGCGGCGCTCTCCGATCGGCGGCGGAGTCCCGGACATCGCC 312  
 QY 256 ATCTCCGGCTGCGGACACCTGAGCCGTTTGTCTGAGGCGGCGGCGAAGTCCGCGTG 315  
 Db 313 CAGGGGCTGACAGCGCCCTACGAGACCCCGCTGAGGCGGAGGACGACGCTCCACAG 372  
 QY 316 GACGCGGGGTACAGCAAGGACGAACTG---GCCGAGATCACCAGAAAGTGCCTGTGATG 372  
 Db 373 CAGGGCGCGGCGGACGACCGCTGATGTGTGGCTACGCGACGAGACGACGCCCGCTCG 432  
 QY 373 TCGGACCTGCGGAAATCTTCTGATCTGACCGCTACCCCGGCGATAGGAAAGCGCAAG 432  
 Db 433 ATGCGCTGCGGATCGAGGTCTGCGGCGGCGGCGGCTCTGCGGCGGCTGACGAGGTCCGCAAG 492  
 QY 433 GCGGAGAGGACCTGTGCAAGCTCAACCCATCAGGTGTATCATCTACGCCATCCGTAACCTG 492  
 Db 493 GACGAGACCGTCCCTACCTGTGCGGCCCGGCGGCAAGACCGAGTACCATGTGATACAG 552  
 QY 493 TGGGCTTCCGCGCGCGCGGAGGATCTTTCGACACACCGCGATGTCGCGGCGCATGTC 552  
 Db 553 GCGAGCGCGCGGCTGTGCGGCTGTGACACCGCTGTCTCTCTCCACGCGCGCGCATTC 612  
 QY 553 CGCGCGCGCGCGCGGCGGCGGAGGATCTTTCGACACCGCGATGTCGCGGCGCATGTC 612  
 Db 613 GACCTGCGGCT 672  
 QY 613 ACCGCGAGCTTTCGAAAGGCGGAGGACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 672  
 Db 673 GCACTGCGGCGGAGGAGGAGGATCAAGCTCGAGAGGAGCACTACCGCTCTCTCTCTCTCT 732  
 QY 673 GACAACTGCTGCGGCGGAGGAGGATCGGCGGCTTCAAGTGTGATCTGCAAGAGCGCAAGCTG 732  
 Db 733 ACCGCGCGGCTTGTGAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 792  
 QY 733 GCTCCCGGCTCCGCGGAGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 792  
 Db 793 ATCATCTGACACTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 852  
 QY 793 GACGAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 852  
 Db 853 CCGTCAAGGTGACCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 912  
 QY 853 GACGAGTGTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 912  
 Db 913 GCGG 972  
 QY 913 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 935  
 Db 973 CCGG 995

RESULT 11  
 ABK91617  
 ID ABK91617 standard; DNA; 9169 BP.

XX ABK91617;  
 AC  
 XX

14-AUG-2002 (first entry)

XX Modified HIV protein-encoding plasmid DNA #169.

XX HIV; human immunodeficiency virus; gene; ds; circular; cyclic; anti-HIV;  
 KW Env; Gag; Pol; Nef; genetic immunisation; protein immunisation; AIDS;  
 KW acquired immunodeficiency syndrome; cytotoxic T lymphocyte; CTL.  
 OS Synthetic.  
 XX





QY 971 GGGCGCTGATGACAGCCGCGCTGATGCAAAACGATCAATCTG 1019  
 DB 8605 ACGGTACACAGCCGCTGCTGCTCAACAGCCGCGGACCG 8653  
 RESULT 14  
 ID AA064201 standard; cDNA; 5392 BP.  
 AC AA064201;  
 XX  
 DT 18-NOV-1994 (first entry)  
 DE Sequence comprising the snab, snab and snac gene cluster.  
 DE Antibiotic; streptomycin; snab; snab; snac; biosynthesis; enzyme;  
 KM biosynthetic pathway; Streptomyces pristinaespiralis; ds.  
 XX Streptomyces pristinaespiralis.  
 OS  
 XX PR696189-A.  
 PN  
 XX PD 01-APR-1994.  
 XX PF 25-SEP-1992; 92FR-0011441.  
 XX PR 25-SEP-1992; 92FR-0011441.  
 XX (RHON) RHONE POULENC RORER SA.  
 PA  
 XX Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P;  
 PI Tibbault D, Zagorec M;  
 DR WPI; 1994-128286/16.  
 XX  
 PT DNA involved in streptomycin antibiotic biosynthesis - for  
 PT prodn. or bio-conversion of streptomycin(s) or prodn. of  
 PT streptomycin intermediates, derivs. or hybrid antibiotics  
 XX  
 PS Disclosure; Page 44-47; 83pp; French.  
 XX  
 CC This sequence comprises the snab, snab and snac genes which are  
 CC involved in the biosynthesis of streptomycin, antibiotics active  
 CC against Gram-positive bacteria. The identification of the sequences  
 CC encoding the enzymes involved in the biosynthetic pathway means that  
 CC they can be isolated and manipulated. Mutant microorganisms in  
 CC which a step in the streptomycin biosynthetic pathway is blocked  
 CC can be cultured to produce streptomycin intermediates, which may  
 CC later be converted to streptomycin derivatives. Recombinant cells  
 CC may also be used for the bioconversion of streptomycin from one  
 CC form to another or for the production of hybrid antibiotics.  
 CC  
 SO Sequence 5392 BP; 811 A; 2161 C; 1671 G; 749 T; 0 other;  
 XX  
 Query Match 5.68; Score 57.4; DB 15; Length 5392;  
 Best Local Similarity 42.18; Pred. No. 0.03;  
 Matches 389; Conservative 0; Mismatches 531; Indels 3; Gaps 1;  
 QY 16 GACCTCGACACCTTCAGCTGTGCGAGCCGCGGCGGAGACACCCG 75  
 DB 3630 GACACCTCTCTCGACGCTCTGCGGAGACCCGCTCAGCGGCGGAGAC 3689  
 QY 76 GCCGCTCGGTGATCAGTACAGGAGTACGAACTGACACCTCCGCTCCGCG 135  
 DB 3690 CTGATCACCACCGGCGAGTCCACATGCGGAGGTCCACCAAGGCTACGCGCC 3749  
 QY 136 GCGCTCGCGGTGATGAGGCGGAGTACCTGCGGCGGAGAGGAGATCTTC 195  
 DB 3750 ATCCGCCCACTGTGCGGAGAGATCTGCGGCGGAGATCTGCTGCGGAGAGG 3809  
 QY 196 GACACCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 255  
 DB 3810 TTGACGCGGCGCTCTGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3869

QY 256 ATCTTCGCGCTGCGGACACCTGACCGGTTGCTGACAGGCGGCTCCAACTGCGCTG 315  
 DB 3870 CAGGCGCTGACAGCGCTGACGAGACCGCGCTGAGGCGGAGAGACGACCTGACAG 3929  
 QY 316 GACGCGGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372  
 DB 3930 CAGGCGGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3989  
 QY 373 TCGAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 432  
 DB 3990 ATGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4049  
 QY 433 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492  
 DB 4050 GACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4109  
 QY 493 TGGGCGCTTCCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552  
 DB 4110 GCGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4169  
 QY 553 CG 612  
 DB 4170 GACGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4229  
 QY 613 ACCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 672  
 DB 4230 GCAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4289  
 QY 673 GACAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 732  
 DB 4290 ACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4349  
 QY 733 GCTCCCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 792  
 DB 4350 ATCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 4409  
 QY 793 GACGAG 852  
 DB 4410 CCGTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4469  
 QY 853 GACGAG 912  
 DB 4470 GCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4529  
 QY 913 GAGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935  
 DB 4530 CCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4552  
 RESULT 15  
 ID AA006844 standard; DNA; 2271 BP.  
 XX  
 AC AA006844;  
 XX  
 DT 05-MAR-1991 (first entry)  
 DE Amylase gene from Streptomyces griseus IMRU 3570.  
 DE Secondary metabolism activation factor gene; amylase gene;  
 KM heterologous protein expression; promoter; SAF; ss.  
 XX Streptomyces griseus.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 318..2155  
 FT /tag=a  
 FT /product= amylase  
 FT /note="including leader sequence"  
 XX  
 MO9014426-A.

PD 29-NOV-1990.  
 XX 18-MAY-1990; 90WO-GB00781.  
 XX 22-SEP-1989; 89US-0410706.  
 PR 19-MAY-1989; 89US-0354265.  
 XX (ISTE) ARS APPL RES SYST.  
 PA Ortega AD, Gil JA, Garcia TV, Martin JF;  
 XX WPI, 1990-375993/50.  
 DR P-PSDB; AAR08263.  
 XX  
 PT Secondary metabolism activation factor gene - isolated from  
 PT Streptomyces and used to increase prodn. of extracellular enzymes  
 PT or heterologous proteins  
 XX  
 PS Disclosure; Fig 11; 61pp; English.  
 XX  
 CC This endogenous gene, with a suitable recognition site BstEII, is  
 CC used for fusion with a foreign gene. The amylase gene encodes  
 CC an extracellular enzyme ensuring secretion of the foreign  
 CC polypeptide or protein. The combination of the SAF gene (AAQ06843)  
 CC with the amylase gene on a plasmid does not cause increase prodn.  
 CC of amylase. Thus, it is preferred to place the foreign protein  
 CC operatively linked to the secretion signal sequence of an  
 CC endogenous extracellular enzyme which is controllable by SAF and,  
 CC pref. further operatively linked to the SAF promoter (in the absence  
 CC of the native promoter), into the chromosomal DNA as opposed to a  
 CC plasmid.  
 CC See also AAQ06843.  
 CC  
 XX  
 SQ Sequence 2271 BP; 358 A; 946 C; 686 G; 281 T; 0 other;

Query Match 5.6%; Score 57.2; DB 11; Length 2271;  
 Best Local Similarity 44.4%; Pred. No. 0.032;  
 Matches 319; Conservative 0; Mismatches 393; Indels 6; Gaps 2;

QY 18 CCTGGGACCTCCAACTGCTGGCCGTCGAGCCCGGCGCATCCGCGAGGACACCCCGGC 77  
 DB 635 CCAGCGGCGCGCGGCTCAAGGTCGTGCGCGACTGCTCATCACACATGGCCGGGGTTC 694  
 QY 78 CGGCTGGGTATCCAGTACAGCGACTAGAGAACTGGACACTCCAGCCGTTCCGGCGG 137  
 DB 695 CGGACCGCGCACCGCGCGACGCGTCAAGAACTAGACTACCGGGCATCTGTCCGG 754  
 QY 138 CGTGGCCGTGATCGAGGCGCAATATCTGCGCGCGCAAGAAAGCAAGATCTCCATCTCGA 197  
 DB 755 CGCGGACATGACACATCTGCGCGCGACGAGATCAACGACTACGCGCAACCGCGCAAGTCCA 814  
 QY 198 CACCGGATTCGGTATTCGATCTGACCTACACCGTCCGCGCATGTATGGACGCGCAACAT 257  
 DB 815 GAAC---TGGGAACTGTGCGGCTCGCGCGACCTGACACCGGTGAGCCGTACGTCCGCA 871  
 QY 258 CTTCCGGCTGGCGGACCACTGAGCCGCTGTCTGAGGCGGCGTCCAAAGCTGGCCTGGA 317  
 DB 872 CGGATCGCGCGCTCACTCAAGACCTGTCTTCTGCTGCGGTGAGCGCTTCGCGCATCGA 931  
 QY 318 CGCGGCGGTACAGAGCAAGCACTGCGCGAGATCAACCAAGAGTGTGATGTGCA 377  
 DB 932 CGCGGCGCAAGACATGCGCCGCG---GCCGACCTACCGCGCATCAAGGCGCAAGGCTCGCA 988  
 QY 378 GCTCGCGGAAATCGTCTGATCTGACCTGACCGCGGAGTACGGAAGAGCGCAAGGCGCA 437  
 DB 989 CGGAGAGCAGTACGAGAGCGAGGCGCATCGAGCGCGGAGGCGCGTCCAGGCCAG 1048  
 QY 438 GAAAGACCTGTCCAAAGCTCAACCATCAAGTGTATCATACCCATCCGTAACCTGTGGGC 497  
 DB 1049 CGAGTACTCTGGCACCGCGCACTGTCAGAGTTCGCGTACCGCGCGCACTCAAGCGGAT 1108  
 QY 498 CTTCCCGCGCGCGAGAGATCTTCCGCGACCAACCGCGCATGTGCGCGCATGTCCGCCG 557

DB 1109 CTTCCAGAACAGAGAACCTGGCCACCCTGAAGAACTTGGCGAGAGACTGGGCTACATGCG 1168  
 QY 558 CGCGGCGCGGACACACCGGTGAGCCCGACCATCAAGAACTACCGTGGGGTATCTCACCCG 617  
 DB 1169 GAGCGGCAAGTCCGCGCTTCTGTGACAAACACGACACGAGCGGGCGCGACACCTT 1228  
 QY 618 AGCCAGTTTCGAAGCCAGAGACCGGTGCGCGCACCGCGGATCTGCTGACTCGACAA 677  
 DB 1229 CAACCTACAAAGAACGGCTCCGCGCTACACCGTCCGCGCGCGTCTTCAATGCTGGCGCTTA 1288  
 QY 678 CTCGCTGGCGCGAAGGTCGCGGCTTCAACGTGTGATCTCAAGAGACGCGCAACCTGCC 735  
 DB 1289 CGGCTCCCGGACGTCACCTCCGCTACGAGTTCCACCGACACGACCGCGCGCGCC 1346

Search completed: April 24, 2003, 19:58:40  
 Job time : 336 secs

GenCore version 5.1.4\_p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:54:42 ; Search time 1541 Seconds  
(without alignments)  
10719.920 Million cell updates/sec

Title: US-10-067-291-2

Perfect score: 1020  
Sequence: 1 atgactgccttcacgacct.....tcgaacgacgcgaatactga 1020

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estlom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	93.6	9.2	657 13	BI997754 1031051G0
2	92	9.0	662 17	BI990060 ATXOA377R
3	82.4	8.1	587 17	BI990170 ATXOD39TF
4	59.8	5.9	925 17	CNS0091P
5	58.8	5.8	433 17	BI990123 ATXOA70TF
6	58	5.7	550 17	AZ935431 BJ_Ba000

7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
58	57	56.4	55.5	55.4	54.6	54.6	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	54.4	
775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775	775
AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	AZ933606	
BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	BI_Ba000	

## ALIGNMENTS

RESULT 1  
BI997754  
LOCUS 1031051G01.Y2 C. reinhardtii CC-1690, Stress II (normalized),  
DEFINITION Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BI997754  
VERSION BI997754.1 GI:1643528  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales.  
REFERENCE 1 (bases 1 to 657)  
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1031  
JOURNAL Unpublished (2001)  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1..657  
/organism="Chlamydomonas reinhardtii"



/strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, Stress II (normalized), lambda Zap II"  
 /note="Vector: phluescript II SK-, site.1: EcoRI, site.2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH<sub>4</sub> - containing) and shifted to TAP - NO<sub>3</sub> - (24hrs); H<sub>2</sub> production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant Phys. 122: 127-135; TAP + H<sub>2</sub>O<sub>2</sub> (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Phluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 116 a 220 c 231 g 90 t  
 ORIGIN

Query Match 9.2%; Score 93.6; DB 13; Length 657;  
 Best Local Similarity 50.0%; Pred. No. 2.3e-08;  
 Matches 262; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

449 CCAAGCTCACCACATCAGGTATCATCTACGCTACCTGAGGCTTCCGCCG 508  
 133 CCAACACACCATCGGCAAGCCCAACATCGTATCGCCGATACAGAGCCCTCCG 192  
 509 CCGAGCAGATTTGGGACCAACCGGATCTGCGCCGATCTCCGCCGCCGCGCA 568  
 193 CGGTGCCAGGAGGAGGCGGATAGGCTGTATCAGGTGACGTGCGCGCGCGCCG 252  
 569 ACACCGTGACCCGACCATCAAGAACTACAGTGGGATGATACCGGACGCTTTCG 628  
 253 ACGTGCAAGACCCGGATGATGATCTCCATCCAAAGTCAATGATGATGCTGCAATCC 312  
 629 AAGCCAAAGACCGTGTGCGGCAACCGGATCTGCTGATCGAGACACTGCTGCGCG 688  
 313 AGCCCAACAGGCGGCTGTGAGAGGCTGTGATGATGAGACCGGCGCGCGCGCA 372  
 689 AAGTCCGGGCTTCAAGTGTGATCTGCT---CAAGAGACGCAAGTGGCTCCCGCTCC 745  
 373 CCTGCAACAGCACTTCTTCTGCTGCGCAAGGCGAGGTGTGGCGCCCGACGCCG 432  
 746 GAACGCGCTTCCGGGATCAACCGGTAAAGCGGTTCGAACCTGGCCGACAGATGGCA 805  
 433 GGCACACAGCTGCGGCGATCACCGGCGGCGGTGCTCAGCTGTGCGCAGCAGCGCA 492  
 806 TCGAAGCCACCTGCGGAGCTCAGACCGGTGAATCTAGACGCGCAGAGTTGATG 865  
 493 TCCCTGCGGTGAGACGAGACTTACTCAAGGAGGTGTAAGGCGGAGAGGCAATTCG 552  
 866 CGGTACACACCGGCGGCGGTGACACCGATCAACTGCTGTGAGGAGGCGCGTGGCA 925  
 553 TGAATGACCTTCTGCGGCGGTCTATCCGCTGTGAGGAGGCGCGGCGGTATAGGCG 612  
 926 ACAGGAGCCCGGTCACTGAGCGGTGCGCAATCCGCGAGCGGTTTC 969  
 613 GCGGACAGCGCGGCGCGCTGCGGAGCGGCTGACAGCTGTAC 656

RESULT 2  
 BH190060  
 LOCUS BH190060 662 bp DNA linear GSS 19-OCT-2001  
 DEFINITION ATXOA37R ATXO Arabidopsis thaliana genomic clone ATXOA37, DNA  
 sequence.  
 ACCESSION BH190060  
 VERSION BH190060.1 GI:16302842  
 KEYWORDS GSS.  
 SOURCE thale cress.

## ORGANISM

Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## CONTACT

## TIGR

## 9712

## Medical Center Drive, Rockville, MD 20850, USA.

## Tel: 301-838-3523

## Fax: 301-838-0208

## Email: cdw@tigr.org

## From Wash. U contig 1142. Caution: the DNA in this BAC may be from

## some non-Arabidopsis source

## Seq primer: TR

## Class: sheared ends.

## Location/Qualifiers

## 1. 662

## /organism="Arabidopsis thaliana"

## /strain="Columbia"

## /db\_xref="taxon:3702"

## /clone\_lib="ATXOA37"

## /note="Vector: PHOS2; site.1: BstXI; 2-3 kb sheared BAC

## DNA inserted into phos2 using BstXI linkers."

## BASE COUNT 134 a 197 c 207 g 124 t

Query Match 9.0%; Score 92; DB 17; Length 662;  
 Best Local Similarity 49.1%; Pred. No. 4.7e-08;  
 Matches 276; Conservative 0; Mismatches 280; Indels 6; Gaps 1;

449 CCAAGCTCACCACATCAGGTATCATCTACGCTACCTGAGGCTTCCGCCG 508  
 2 CCTCTCTGTGACATCGGCGAGCATGTCATCATCTGACATCAACATCACTCTATCCGA 61  
 509 CCGAGCAGATTTGGGACCAACCGGATCTGCGCCGATCTCCGCCGCCGCGCA 568  
 62 AGAGCAGATCTTGGGACCAACCGGATCTGCGCCGATCTCCGCCGCCGCGCA 121  
 569 ACACCGTGACCCGACCATCAAGAACTACAGTGGGATGATACCGGACGCTTTCG 628  
 122 CGCGGTGATGCGGAGGAGGATGCTGATGATGATGATGATGATGATGATGATGATG 181  
 629 AAGCCAAAGACCGTGTGCGGCAACCGGATCTGCTGATCGAGACACTGCTGCGCG 688  
 182 AAGCAGATCCAGCGGCTGTGAGAGGCTGTGATGATGATGATGATGATGATGATG 241  
 689 AAGTCCGGGCTTCAAGTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 748  
 242 AATGACGCGGCGCAATATTTCTGATCGGATGATGATGATGATGATGATGATGATG 301  
 749 -----ACGCTTCCGCGCATCACCGGTAAAGCGGTTCGAACCTGGCCGACGATG 802  
 302 GCGAGCGCGCGTGGAGGACATCACCGCGGCGGTGATGATGATGATGATGATGATG 361  
 803 GCATGCAACACCTGCGGCGGAGCTCAGACCGGTGAATCTAGACGCGCAGAGTTGA 862  
 362 GATGCGGATGAGAGCGGCGGATGAGCGGCTATGATGATGATGATGATGATGATG 421  
 863 TGGCGGTACACCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATG 922  
 422 TCTTACGCGGCGACCGCGGCGGAGGTATTCGCTGTGATGATGATGATGATGATGATG 481  
 923 GCAAGCGGCGAGCCCGGTCACTGAGCGGTGCGCAATCCGCGAGCGGTTTCGCGTGTATG 982  
 482 GCGATGCAAGCGCGGCTGCTTACGAGAGGCTTATGATGATGATGATGATGATGATG 541  
 983 ACGAGCGGCGCGCGCTGATGCA 1004







LOCUS A2933606 775 bp DNA linear GSS 24-Apr-2001  
 DEFINITION Bt\_Ba001E09f B. japonicum BAC library Bradyrhizobium japonicum  
 ACCESSION A2933606  
 VERSION A2933606.1 GI:13775666  
 KEYWORDS GSS  
 SOURCE Bradyrhizobium japonicum.  
 ORGANISM Bradyrhizobium japonicum; Alphaproteobacteria; Rhizobiales;  
 Bacteria; Proteobacteria; Bradyrhizobium.  
 REFERENCE 1 (bases 1 to 775)  
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Golcochea  
 J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome  
 TITLE genome Res. 11 (8), 1434-1440 (2001)  
 JOURNAL  
 MEDLINE 21376150  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Class: BAC ends  
 High quality sequence stop: 716.  
 Location/Qualifiers  
 source 1..775  
 /organism="Bradyrhizobium japonicum"  
 /strain="USD110"  
 /db\_xref="taxon:375"  
 /clone\_1lb="B. japonicum BAC library"  
 /lab\_host="E. coli"  
 /note="Vector: pindig0536; site\_1: HindIII"  
 BASE COUNT 135 a 241 c 233 g 164 t  
 ORIGIN  
 Query Match 5.7%; Score 58; DB 17; Length 775;  
 Best Local Similarity 51.1%; Pred. No. 0.17;  
 Matches 136; Conservative 0; Mismatches 130; Indels 0; Gaps 0;  
 QY 662 TCTCTGACTGGACACAGTGGCCGAGGCTTCAGCTGATGCTCAAG 721  
 Db 139 TTCCGGACCTGCTCTCAGACAGCGCGCTTCAACGTAACCTGCGAGCGCA 198  
 QY 722 ACGGACGTGGCTCCCGCTCCGGAACGCGTTCGCGGATCAACCGTGA 781  
 Db 199 AGCTTACGTGGCTCCGAGATTCAGACTGCTGGTGTCAACGGCAAGCGCG 258  
 QY 782 TCGAAGTGGCGACAGATGGGATCGAAGCCACCTGCGGAGCTACACCGCTGA 841  
 Db 259 TCAATATCGGATCCAGAGTGGCGCAACGCCAATCTTCGAGCTATCAAGGCTCC 318  
 QY 842 TCTAGACGCGACGAGTGTGATGGGCTACACCGCGGGGCTACACCGATCACT 901  
 Db 319 GCGCGCTATATCCGACATCAAGGCGCAGACGCGAGGGCTGATCCGAGTATCT 378  
 QY 902 CGCTGATGGCGAGCGCGTGGGCAAC 927  
 Db 379 ACGACTTCAACGATTTCTGTAACAGC 404  
 RESULT 8  
 A2934707 582 bp DNA linear GSS 24-Apr-2001  
 LOCUS A2934707  
 DEFINITION Bt\_Ba0002D18r B. japonicum BAC library Bradyrhizobium japonicum  
 ACCESSION A2934707  
 VERSION A2934707.1 GI:13776767  
 KEYWORDS GSS.  
 SOURCE Bradyrhizobium japonicum.  
 ORGANISM Bradyrhizobium japonicum; Alphaproteobacteria; Rhizobiales;  
 Bacteria; Proteobacteria; Bradyrhizobium.

REFERENCE 1 (bases 1 to 582)  
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Golcochea  
 J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome  
 TITLE genome Res. 11 (8), 1434-1440 (2001)  
 JOURNAL  
 MEDLINE 21376150  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Class: BAC ends  
 High quality sequence start: 20  
 High quality sequence stop: 548.  
 Location/Qualifiers  
 FEATURES  
 source 1..582  
 /organism="Bradyrhizobium japonicum"  
 /strain="USD110"  
 /db\_xref="taxon:375"  
 /clone\_1lb="B. japonicum BAC library"  
 /lab\_host="E. coli"  
 /note="Vector: pindig0536; site\_1: HindIII"  
 BASE COUNT 114 a 186 c 159 g 123 t  
 ORIGIN  
 Query Match 5.5%; Score 56.4; DB 17; Length 582;  
 Best Local Similarity 50.8%; Pred. No. 0.32;  
 Matches 135; Conservative 0; Mismatches 131; Indels 0; Gaps 0;  
 QY 662 TCTCTGACTGGACACAGTGGCCGAGGCTTCAGCTGATGCTCAAG 721  
 Db 179 TTCCGGACCTGCTCTCAGACAGCGCGCTTCAACGTAACCTGCGAGCGCA 238  
 QY 722 ACGGACGTGGCTCCCGCTCCGGAACGCGTTCGCGGATCAACCGTGA 781  
 Db 239 ACGTACCTGGGCTCCGAGATTCAGACTGCTGGTGTCAACGGCAAGCGCG 298  
 QY 782 TCGAAGTGGCGACAGTGGGATCGAAGCCACCTGCGGAGCTACACCGCTGA 841  
 Db 299 TCTATATGGGATCCAGATGGGCGCAACGCCAATCTTCGAGCTATCAAGGCTCC 358  
 QY 842 TCTAGACGCGACGAGTGTGATGGGCTACACCGCGGGGCTACACCGATCACT 901  
 Db 359 GCGCGCTATATCCGACATCAAGGCGCAGACCGGAGGGCTGATCCGAGTATCT 418  
 QY 902 CGCTGATGGCGAGCGCGTGGGCAAC 927  
 Db 419 ACGACTCCACGATTTCTGTAACAGC 444  
 RESULT 9  
 A2933996 690 bp DNA linear GSS 24-Apr-2001  
 LOCUS A2933996  
 DEFINITION Bt\_Ba0001L03f B. japonicum BAC library Bradyrhizobium japonicum  
 ACCESSION A2933996  
 VERSION A2933996.1 GI:13776056  
 KEYWORDS GSS.  
 SOURCE Bradyrhizobium japonicum.  
 ORGANISM Bradyrhizobium japonicum; Alphaproteobacteria; Rhizobiales;  
 Bacteria; Proteobacteria; Bradyrhizobium.  
 REFERENCE 1 (bases 1 to 690)  
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Golcochea  
 J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome  
 TITLE genome Res. 11 (8), 1434-1440 (2001)  
 JOURNAL  
 MEDLINE 21376150

COMMENT Contact: Ming RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: twang@clemson.edu

FEATURES  
source  
1. 690  
/organism="Bradyrhizobium japonicum"  
/strain="USDA110"  
/db\_xref="taxon:375"  
/clone\_lib="B. japonicum BAC library"  
/lab\_host="E. coli"  
/note="Vector: pIndigo536; Site\_1: HindIII"

BASE COUNT 125 a 226 c 193 g 145 t 1 others

ORIGIN

Query Match 5.5%; Score 56.4; DB 17; Length 690;  
Best Local Similarity 50.8%; Pred. No. 0.33;  
Matches 135; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 662 TGCCTGACCTGGACAACTGGCTGGCCGAGGCTCCGCTTCAACGTGTGCTCAAG 721  
DB 139 TTCCGACCTGCTGCTCAACGAGCCGCGCTTCAACGTCAAGCTGCCGAGTGGCA 198

QY 722 ACGCAGCTGGGCTCCGCTCCGGAAGGCTTGGCGGCAATACCCGTAACAGCTGT 781  
DB 199 ACGTTAGCTGGCTCCGAGGATTAAGCATCTGTGCTTCAACGAGGACCGGCGG 258

QY 782 TCGAATGCGCCGACAGATGGGCAATCGAACCACTGCGGAGCTCACCAGCGCTGAAC 841  
DB 259 TCTATATCGGCAATCGAGTGGCGCAACCACTTTCGACGTCATCAAGAGGCTGC 318

QY 842 TCTACGAGCGCGGAGGATGAGCGGTTCACACCCGCGGCGGGGTTCACACGATCACT 901  
DB 319 GCGCCGCTTATCCCGAATCAAGGCGGAGGAGCGGAGCGGAGGCTGAACTCGAGATCATCT 378

QY 902 CGCTGATGTCGAGCGCGTGGGCAAC 927  
DB 379 ACGACTTCACCGATTCGTGAACAGC 404

RESULT 10  
BIT20731 488 bp mRNA linear EST 19-SEP-2001  
LOCUS  
DEFINITION 1031051G01.Y1 C. reinhardtii CC-1690, Stress II (normalized),  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION BIT20731  
VERSION BIT20731.1 GI:15696426  
KEYWORDS  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales; Chlamydomonadales;  
1 (bases 1 to 488)

REFERENCE  
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,  
P., McDermott, J. P., Shlager, J., Silflow, C. and Stern, D.,  
Analyses of the Chlamydomonas reinhardtii genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1031  
Unpublished (2001)

JOURNAL  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu,  
Location/Qualifiers

FEATURES

source  
1. 488  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress II (normalized)  
, Lambda Zap II"  
/note="Vector: pluescript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Stress condition II library, constructed by John  
Davies and Jeffrey McDermott, combines cDNAs from CC-1690  
cells grown to mid-log phase in TAP (NH4+ - containing)  
and shifted to TAP - NO3- (24hrs) H2 production  
conditions (0, 12hr, 24hr) see Wells et al., (2000) Plant  
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +  
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).  
Polya mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
zap II (Stratagene) in the EcoRI (5') and XhoI (3')  
sites. Pluescript II SK- plasmids were excised from the  
lambda zap clones by superinfection with Exsist  
(Stratagene) phage. The library was normalized using  
method 4 described in Bonaldo et al., (1996) Genome  
Research 6: 791-806."

BASE COUNT 92 a 167 c 165 g 64 t

ORIGIN

Query Match 5.4%; Score 55.4; DB 13; Length 488;  
Best Local Similarity 49.3%; Pred. No. 0.49;  
Matches 174; Conservative 0; Mismatches 176; Indels 3; Gaps 1;

QY 449 CCAAGCTCACCCATGAGTGTACATCTAGCCATCCCTGACCTGGGCTTCCGCGCCG 508  
DB 133 CCAACACCACTATGCGGCAAGCCACCATCTGTAATGCTCCGAGTCAAGAGGCTTGC 192

QY 509 CCGACCAATCTTTCGGGACACCGGATGTCGCGGATATGTCGCGCGCGCGCA 568  
DB 193 CGGTCCGAGGAGGAGGCGGCAATAGGCTTTCACGCTGACGCTGCGGCTGCGCGCG 252

QY 569 ACACCTGACCCGACATCAAGAACTACAGTGGGGTGATCTCAACGCGGCACTTTTCG 628  
DB 253 ACGTCAAGAGACCCCGGATGAGAACTCCCATCAAGCTCAATGCTGCTGCTATCC 312

QY 629 AAGCCAGAGACCGTGTGCGGACCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 688  
DB 313 AGGCCAACAAGCGGCGGTGTGACGAGGCTGTGATCTGATGAGACCCGAGGCTGTGGCCA 372

QY 689 AAGTCCGCGCTCAAGTGTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 745  
DB 373 CCTGCAACAGCACTTCTTCAATCTGCGGAGGCGAGGTGTGGGCGCCGAGCCGCG 432

QY 746 GGAACGCGTTCGCGGACATCACCGGTAAGAGCGTTCGAACTGCGGACGACG 798  
DB 433 GGCACACAGCTGCGGCGATCACCGGCGGCGGTGCTCAGCTGTGCGACAGC 485

RESULT 11  
AV633533 478 bp mRNA linear EST 15-DEC-2000  
LOCUS  
DEFINITION AV633533 Chlamydomonas reinhardtii 5% CO2 chlamydomonas reinhardtii  
cDNA clone HC021h07\_r 5', mRNA sequence.  
ACCESSION AV633533  
VERSION AV633533.1 GI:10776853  
KEYWORDS  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonadales;  
1 (bases 1 to 478)

REFERENCE  
AUTHORS Asamizu, E., Miura, K., Kuchio, K., Inoue, Y., Fukuzawa, H., Ohnaka, K.,  
Nakamura, Y. and Tabata, S.,  
Generation of expressed sequence tags from low-CO2 and high-CO2  
adapted cells of Chlamydomonas reinhardtii  
DNA Res. 7 (5), 305-307 (2000)

JOURNAL  
COMMENT 20539644

FEATURES



VERSION B0812089.1 GI:22055862  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii.  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 607)  
 AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shlager, J., Sillfow, C. and Stern, D.  
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1030  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Charles Hauser  
 DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000  
 Tel: 919 613 8159  
 Fax: 919 613 8177  
 Email: chauser@duke.edu.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..607  
 /organism="Chlamydomonas reinhardtii"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, Deflagellation (normalized), Lambda zap II"  
 /note="Vector: Bluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Deflagellation library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells which had been re-synthesizing flagella for 15, 30 and 60 min after being deflagellated by pH shock. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda zap clones by superinfection with Exassist (Stratagene) phage. The library was normalized using method 4 described in Bonaldi et al., (1996) Genome Research 6: 791-806."  
 BASE COUNT 115 a 202 c 183 g 107 t  
 ORIGIN  
 Query Match 5.4%; Score 54.6; DB 14; Length 607;  
 Best Local Similarity 47.0%; Pred. No. 0.72;  
 Matches 214; Conservative 0; Mismatches 229; Indels 12; Gaps 1;  
 QY 554 GCCGCGCGCGCGCAACCGTCGACCGCAGTCAAGACTACAGTGGGTGATCTCA 613  
 DB 163 GCCGCTCTCGCGCGCGCTGTGTGCGCGCGCAAGTAGAGTCAATTAAGA 222  
 QY 614 CCGGAGGCAATTTCAGAACCCAGAGACCGTGTGCGCGCACCGGATCTGCTGACTCG 673  
 DB 223 CCGGTGGCAGCGTTCCTCCCAAGCGCGCGGT-----ATCTGCGCCATGAGCG 270  
 QY 674 ACAACTGCTGGCGGAGAGTCCGGGCTTCAAGTGTGATCTGATCGAAGAGCGGACCTGG 733  
 DB 271 AGTCAACGGCAGCTGGGGAAGCGCTGACATCTGCGCGTGGGAACACCGAGAGAGA 330  
 QY 734 CTTCCCGCTCCGGAACGCGTGGCGGCAATCCGTTAAGAGCGTGTGCAATCGGCGG 793  
 DB 331 ACCGCGCGCGCTACCGCGAGACTGTGTGACGCGCGCGCGCGCGCGCATCATCTCGG 390  
 QY 794 ACCAGATGGGATGAGAGCGACCGTCCGGGAGCGTCAACAGCCGTGAATCTGACAGCGCG 853  
 DB 391 GCGCTATCTCTTTCGAGAGACCGCTTATCAATGTCACCGCTCCCGCAAGAGTTCTGG 450  
 QY 854 ACAGATGTAGTGGCGGTCCACCGCGCGGGGTCAACCGATCAACTCTGAGATGGCG 913  
 DB 451 ATGTGATGAAGAGAGCAAGATCTGTCGCCGCGCATCAAGAGAGGCTGTGGTCCCG 510  
 QY 914 AGCGCTGGGCAAGCGCGCGGTCTCACTAGCGTGGTCCATCCGCGACCGTGTGGG 973  
 DB 511 TGTCCAAACCAACGAGTGTGATGAGCGCGCTTGAAGCGCGCTTGAAGAGCGCTGGG 570

QY 974 CGCTGATGACGAGCGCGCGCGCTGATCGAAGC 1008  
 DB 571 CTGAGTACTACAAGCGCGCGCTCGCTTCCGCAAG 605  
 RESULT 14  
 AM757472  
 LOCUS 621 bp mRNA linear EST 03-MAY-2000  
 DEFINITION 67401D07.y1 C. reinhardtii CC-1690, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
 ACCESSION AM757472  
 VERSION AM757472  
 KEYWORDS AM757472.1 GI:7686824  
 SOURCE EST.  
 ORGANISM Chlamydomonas reinhardtii.  
 Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 621)  
 AUTHORS Grossman, A., Davies, J., Federpiel, N., Harris, E., Lefebvre, P., McDermott, J.P., Sillfow, C., Stern, D. and Surzycki, R.  
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Elizabeth H. Harris  
 DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000, USA  
 Tel: 919 613 8164  
 Fax: 919 613 8177  
 Email: chlamy@duke.edu.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..621  
 /organism="Chlamydomonas reinhardtii"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, Lambda zap II"  
 /note="Vector: bluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda zap clones by superinfection with Exassist (Stratagene) phage."  
 BASE COUNT 111 a 215 c 188 g 107 t  
 ORIGIN  
 Query Match 5.4%; Score 54.6; DB 10; Length 621;  
 Best Local Similarity 47.0%; Pred. No. 0.72;  
 Matches 214; Conservative 0; Mismatches 229; Indels 12; Gaps 1;  
 QY 554 GCCGCGCGCGCGCAACCGTCGACCGCAGTCAAGACTACAGTGGGTGATCTCA 613  
 DB 115 GCCGCTCTCGCGCGCGCTGTGTGCGCGCGCAAGTAGAGTCAATTAAGA 174  
 QY 614 CCGGAGGCAATTTCAGAACCCAGAGACCGTGTGCGCGCACCGGATCTGCTGACTCGG 673  
 DB 175 CCGGTGGCAGCGTTCCTCCCAAGCGCGCGGT-----ATCTGCGCCATGAGCG 222  
 QY 674 ACAACTGCTGGCGGAGAGTCCGGGCTTCAAGTGTGATCTGATCGAAGAGCGGACCTGG 733  
 DB 223 AGTCAACGGCAGCTGGGGAAGCGCTGACATCTGCGCGTGGGAACACCGAGAGAGA 282  
 QY 734 CTTCCCGCTCCGGAACGCGTGGCGGCAATCCGTTAAGAGCGTGTGCAATCGGCGG 793  
 DB 283 ACCGCGCGCGCTACCGCGAGCTGTGTGACCGCGCGCGCGCGCGCATCATCTCGG 342







GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 19:43:06 ; Search time 98 Seconds  
(without alignments)  
3191.942 Million cell updates/sec

Title: US-10-067-291-2

Perfect score: 1020  
Sequence: 1 atgactgtcttcaagacct.....tcgaacgacgataactga 1020

Scoring table: IDENTITY-MTC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :  
1: Issued Patents, NA:\*  
2: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/PCRTUS.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1020	100.0	1020	US-09-527-522-2	Sequence 2, Appl
2	66.2	6.5	15872	US-09-105-537-1	Sequence 1, Appl
3	57.8	5.7	12388	US-08-387-942C-1	Sequence 1, Appl
4	57.4	5.6	1208	US-08-403-852D-4	Sequence 4, Appl
5	57.4	5.6	1208	US-08-510-646B-4	Sequence 4, Appl
6	57.4	5.6	1208	US-09-231-818-4	Sequence 4, Appl
7	57.4	5.6	5392	US-08-403-852D-1	Sequence 1, Appl
8	57.4	5.6	5392	US-08-510-646B-1	Sequence 1, Appl
9	57.4	5.6	5392	US-09-231-818-1	Sequence 1, Appl
10	57	5.6	1140	US-09-023-173-4	Sequence 4, Appl
11	55.4	5.4	1162	US-09-173-300-21	Sequence 21, Appl
12	53.8	5.3	1294	US-09-025-691-2	Sequence 2, Appl
13	53.6	5.3	2634	US-08-196-218-31	Sequence 31, Appl
14	53.6	5.3	2634	US-08-681-953-31	Sequence 31, Appl
15	53.2	5.2	735	US-09-003-287-7	Sequence 7, Appl
16	51.6	5.1	3756	US-08-576-626A-1	Sequence 14, Appl
17	50.4	4.9	1086	US-09-173-300-14	Sequence 2, Appl
18	50.2	4.9	1835	US-09-417-704-2	Sequence 2, Appl
19	50	4.9	20235	US-07-642-734C-3	Sequence 3, Appl
20	50	4.9	403765	US-08-439-009A-3	Sequence 3, Appl
21	50	4.9	403765	US-09-103-840A-2	Sequence 2, Appl
22	48.8	4.8	1693	US-09-320-878-23	Sequence 23, Appl
23	48.2	4.7	1473	US-08-541-033A-25	Sequence 25, Appl
24	48.2	4.7	1473	US-08-828-451-25	Sequence 25, Appl
25	48.2	4.7	1506	US-08-541-033A-23	Sequence 23, Appl
26	48.2	4.7	1506	US-08-828-451-23	Sequence 23, Appl
27	48.2	4.7	1969	US-08-541-033A-7	Sequence 7, Appl

28	48.2	4.7	1969	2	US-08-828-451-7	Sequence 7, Appl
29	48.2	4.7	2096	2	US-08-541-033A-19	Sequence 19, Appl
30	48.2	4.7	2096	2	US-08-828-451-19	Sequence 19, Appl
31	48.2	4.7	2099	2	US-08-541-033A-3	Sequence 3, Appl
32	48.2	4.7	2099	2	US-08-828-451-3	Sequence 3, Appl
33	48.2	4.7	2137	2	US-08-541-033A-18	Sequence 18, Appl
34	48.2	4.7	2137	2	US-08-828-451-18	Sequence 18, Appl
35	48.2	4.7	2140	2	US-08-541-033A-1	Sequence 1, Appl
36	48.2	4.7	2140	2	US-08-828-451-1	Sequence 1, Appl
37	48	4.7	1143	2	US-08-387-942C-19	Sequence 19, Appl
38	47.8	4.7	2064	1	US-08-343-428-1	Sequence 1, Appl
39	47.8	4.7	2167	3	US-08-461-775-9	Sequence 9, Appl
40	47.8	4.7	2167	3	US-09-031-606-9	Sequence 9, Appl
41	47.8	4.7	2668	3	US-08-461-775-11	Sequence 11, Appl
42	47.8	4.7	2668	3	US-09-031-606-11	Sequence 11, Appl
43	47.4	4.6	1288	1	US-08-440-856A-9	Sequence 9, Appl
44	47.2	4.6	8051	2	US-08-576-626A-2	Sequence 2, Appl
45	47	4.6	390	4	US-09-197-649-7	Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-09-527-522-2  
Sequence 2, Application US/09527522  
Patent No. 6413752

GENERAL INFORMATION:  
APPLICANT: Takashima, Yoshiki  
TITLE OF INVENTION: Protein capable of catalyzing transamination  
TITLE OF INVENTION: stereoselectively, gene encoding said protein  
FILE REFERENCE: 058251  
CURRENT APPLICATION NUMBER: US/09/527,522  
CURRENT FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: JP 11/075511  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 11/088634  
PRIOR FILING DATE: 1999-03-30  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1020  
TYPE: DNA  
ORGANISM: Mycobacterium aurum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1017)  
US-09-527-522-2

Query Match 100.0%; Score 1020; DB 4; Length 1020;  
Best local similarity 100.0%; Pred. No. 7.6e-214; Indels 0; Gaps 0;  
Matches 1020; Conservative 0; Mismatches 0

QY	1	ATGACTGCTTTTGAAGCTTGGACCTCCACCTGGTGTGGCCGCGGCGGCATC	60
DB	1	ATGACTGCTTTTGAAGCTTGGACCTCCACCTGGTGTGGCCGCGGCGGCATC	60
QY	61	CGCGAGACACCCGCGCGGCTGCTGATCCAGTACGACGATTCGACACCTTC	120
DB	61	CGCGAGACACCCGCGCGGCTGCTGATCCAGTACGACGATTCGACACCTTC	120
QY	121	AGCCGCTTGGCGGCGGCTGCTGATCCAGTACGACGATTCGACACCTTC	180
DB	121	AGCCGCTTGGCGGCGGCTGCTGATCCAGTACGACGATTCGACACCTTC	180
QY	181	AGATCTCCATCTTGCACACCGATTCGATTCGATTCGATTCGATTCGATTC	240
DB	181	AGATCTCCATCTTGCACACCGATTCGATTCGATTCGATTCGATTCGATTC	240
QY	241	GTATGACACGACGACATCTTCCGCTCGGCGACACCACTGGACCGGTTCG	300
DB	241	GTATGACACGACGACATCTTCCGCTCGGCGACACCACTGGACCGGTTCG	300

```

Db 241 GATGAGCAGGCGACATCTCCGCTCGCGACACACCTGGACCCGCTTGTGACGGGGCG 300
Qy 301 TCACAGCTGCGCCCTGACGCGCGGGTACAGACAGCAAGCAAGCGGCGGATACCAAGAG 360
Db 301 TCACAGCTGCGCCCTGACGCGCGGGTACAGACAGCAAGCAAGCGGCGGATACCAAGAG 360
Qy 361 TCGGTGTGATGTGCGACGCTCGCGAATCGTTCTGATCTGACCGCTACCGGGGATAC 420
Db 361 TCGGTGTGATGTGCGACGCTCGCGAATCGTTCTGATCTGACCGCTACCGGGGATAC 420
Qy 421 GGAAGGCGCAAGGGGAGGAAGGACCTGTCCAGCTACCCATCATGAGTGTACATTCAGCC 480
Db 421 GGAAGGCGCAAGGGGAGGAAGGACCTGTCCAGCTACCCATCATGAGTGTACATTCAGCC 480
Qy 481 ATCCCGTACCTGTGGGCTTCCCGCCCGGAGCAGATCTTGGCAGCACCGCGGATCGT 540
Db 481 ATCCCGTACCTGTGGGCTTCCCGCCCGGAGCAGATCTTGGCAGCACCGCGGATCGT 540
Qy 541 CCGGCGCATGTCCGCGCGCGCGCGCGCGCGAGACACCGCTGACCCGACCGGATCGT 600
Db 541 CCGGCGCATGTCCGCGCGCGCGCGCGCGCGAGACACCGCTGACCCGACCGGATCGT 600
Qy 601 TGGGATGATCTCACCGCAGCAGATTTGAGAGCCAGGACCGTGTGGCGGACCGGATC 660
Db 601 TGGGATGATCTCACCGCAGCAGATTTGAGAGCCAGGACCGTGTGGCGGACCGGATC 660
Qy 661 CTGCTGATCTGCGACAACTCGTGGCGGAGGTCGCGGCTTCAACGTGTGATCTGCAAG 720
Db 661 CTGCTGATCTGCGACAACTCGTGGCGGAGGTCGCGGCTTCAACGTGTGATCTGCAAG 720
Qy 721 GAGGCGAAGGTGGCGCGCGCGCGCGCGCGAGACGCTTGGCGGACCGGATCGT 780
Db 721 GAGGCGAAGGTGGCGCGCGCGCGCGCGCGAGACGCTTGGCGGACCGGATCGT 780
Qy 781 TTGCAACTGCGCAGCAGATGAGTGGATCGAAGCACCCTGCGCAGCTGACACCGCTGAA 840
Db 781 TTGCAACTGCGCAGCAGATGAGTGGATCGAAGCACCCTGCGCAGCTGACACCGCTGAA 840
Qy 841 CTCTAGACGCGCCAGAGTTGATGGCGGTACACCGCGGCGGCTGACACCGGATCAAC 900
Db 841 CTCTAGACGCGCCAGAGTTGATGGCGGTACACCGCGGCGGCTGACACCGGATCAAC 900
Qy 901 TCGCTGATGAGCGAGCGCGTGGGCAACGCGGAGCGCGGCTGCACTGAGGTTGGCCATCCG 960
Db 901 TCGCTGATGAGCGAGCGCGTGGGCAACGCGGAGCGCGGCTGCACTGAGGTTGGCCATCCG 960
Qy 961 GACCGGTTCTGGGCGCTGATGAGCAGCGCGGCGCGCTGATGAGAAAGATGACTGA 1020
Db 961 GACCGGTTCTGGGCGCTGATGAGCAGCGCGGCGCGCTGATGAGAAAGATGACTGA 1020

```

## RESULT 2

```

; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

```

Query Match 6.5%; Score 66.2; DB 4; Length 15872;
Best Local Similarity 44.4%; Pred. No. 7, 8e-06;
Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;
Qy 50 CCGGCGCATCCCGGAGGACACCGCGCGGCTGCGTATCCAGTACAGCACTACGAAAC 109
Db 765 CCGGCGCGCTCCCGGAGGAGACCGGTGTCCAGGCTTATCCGCGGACGCGCTCAACA 824
Qy 110 TGGACACCTTCAGCCGCTTCCCGCGCGGCTGCTGATGAGGAGGAGTACCTGCGG 169
Db 825 ACAGGAGGACACCGCGGCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 884
Qy 170 CCGAAG--AAGGAGATCTCATCTTGTGACACCGAGTGTGATTCATTCGATGACT 226
Db 885 GCGAGGCGTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 944
Qy 227 ACACGCTCGCGATGTATGAGCAGGCA--ACATCTTCCGCTGCGGACACACCTGAGC 283
Db 945 GCACCGAAGCCCGCGGCTGCGGAGCAGCCATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1004
Qy 284 GGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 343
Db 1005 CCGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1064
Qy 344 CCGAGATCACCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 403
Db 1065 TCGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1124
Qy 404 CCGTACACCGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
Db 1125 GATCCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1184
Qy 464 AGGTGATATCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 523
Db 1185 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1244
Qy 524 GCACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 583
Db 1245 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1304
Qy 584 CCAATCAAGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643
Db 1305 CCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1364
Qy 644 GTGCGCGCAGCGCATCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 703
Db 1365 TGCCTTGTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 1424
Qy 704 ACGTGTGATGTGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 763
Db 1425 ACGAGGCGGCTGGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1484
Qy 764 TCACCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 823
Db 1485 CCACCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 884
Qy 824 ACGTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 884
Db 1545 TCGACGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1571

```

## RESULT 3

```

; Sequence 1, Application US/08387942C
; Patent No. 5939289
; GENERAL INFORMATION:
; APPLICANT: ERTESVAG, HELEGA
; APPLICANT: VALIA, SVEIN
; APPLICANT: SKJAK-BRAEK, GUDMUND
; APPLICANT: LARSEN, BJORN
; TITLE OF INVENTION: DNA COMPOUNDS COMPRISING SEQUENCES
; OF INVENTION: ENCODING MANNITOLAN C-5-EPIMERASE
; NUMBER OF SEQUENCES: 52
US-08-387-942C-1

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/387,942C
FILING DATE: 09-MAY-1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1809-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12588 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Azotobacter vinelandii
STRAIN: E

FEATURE:
NAME/KEY: CDS
LOCATION: 290..1951
FEATURE:
NAME/KEY: CDS
LOCATION: 2227..6438
FEATURE:
NAME/KEY: CDS
LOCATION: 6702..9695
FEATURE:
NAME/KEY: CDS
LOCATION: 9973..12588
US-08-387-942C-1

Query Match 5.7%; Score 57.8; DB 2; Length 12588;
Best Local Similarity 43.3%; Pred. No. 0.00051;
Matches 437; Conservative 0; Mismatches 557; Indels 15; Gaps 3

OY 20 TCGGCACCTCCACCTGCTGTCGAGACCCGGGCGCATCCGCGAGACACCCGGCGG 79
Db 7651 TCTAACCCACGACCCGCACTGTGATCGAAGGCACACATCTCGTTCGCACTCCA 7710
OY 80 GCTCGGTATCCAGTACGAGCGCATCGAAGTGAACACTCCAGCCCGTTCGCGGCGG 139
Db 7711 CCTATGGCATCCAGAGGCGCGACGCGACCGACTACGACGCTCTACGCGCAACAGG 7770
OY 140 TCGCCTGATCGAGGCGGAATACCTGCGCGCGCGGAAGAGCAAGATCTTCATCTTGACA 199
Db 7771 TCAGCAATGTGAGAAACGGCTCGGTGCGCCTCTACGCGCGCAACTCCGCTCCGACC 7830
OY 200 CCGGATTTGGGTATTCCTCGATCTGACCTTCACCTGTCGGCGCATGTATGAGCACGGCAACATCT 259
Db 7831 TGCCCCGACCGCGGACGAGCGGACCTCTGAAGGACGCGCG-----GCAACGACACGC 7884
OY 260 TCCGGCTCGCGCACCACTGAGACCGATTGCTCGACGGGCGCTTCAGAGCTGCGCTTGAGAG 319
Db 7885 TTGGGGGACGACGACCGCCACGAGAGCGTCTCGGGCTTGAGCGGCAAGACGCGCTGAAAG 7944
OY 320 CCGGGTACGACGAGCACTGGCCGGAATACCAAGAAATGGTGTGATGTGCGACG 379

```

```

Db 7945 GCGGCGCGCGGCAACGACATCTTCGAGCGGGGGCGCGCGGCGAGCAACCTGACCCGGCGCG 8004
Oy 380 TTGGCGCAATGTTTCGGAATCTGACAGCTGCACCCCGGGGATPACGGAAAGCGCAAGGCGGAGA 439
Db 8005 CGGCGCGCGGCACTGTTTCCGGCTTCCTCGCGCGGACCGACAGACTACCGACACCGACAGCCCCA 8064
Oy 440 AGGACCTGTCCAAAGCTCACACCATCAGGTGTATCATCTAGCCATCCGTAACCTGTGGGCT 499
Db 8065 GCTTCAACGACACTGATATCACCAGCTTGCAGCGCCAGCCAGACCGCATGTGACCTGTGCCGCG 8124
Oy 500 TCCGCGCCCGCGGACAGATCTTGGGACACACCGCGCATGTGCGCGGCATCTCCGCGCGG 559
Db 8125 TTGGGCTTACCGGGCTGGGGGAGCGGTATPAGGGCAACCTGTGTGTGAGGTACGCGCG 8184
Oy 560 CCGGCGCCCAACACCGCTGCAGCCGACCATCAAAACATACACAGTGGGGGTATPACACCGCAG 619
Db 8185 ACGGCGACCGCGCACTATCTGAAAGAGCTTGGAGGGGATGCGAGGGGGGGCGGTTTCGAGA 8244
Oy 620 CCAATTTCGAAGCCAAAGACCGGTGTGCGC---GCAACCGCATCTCTGTGCATCGGACA 676
Db 8245 TTGGCTTCGAGACGGAACTTTCGCGGCTCTCTGTGTCGCGGCAACCTGTCTTTCAGAGCGCA 8304
Oy 677 ACTGCGTGGCGGAAGGTCGCGGGTTCGAAGG-----TGTGCATGCTCAAGGACGGCAAGC 730
Db 8305 CCGGCATTCGAGGGGATGCGGCGGACAAAGCCCTGCTGCTGTACTTCGCGCGCCGACAGAT 8364
Oy 731 TTGGCTTCCCGTCCCGGGAACGCGTTGCTCGGGGATCACCCTGTAAAGAGGTTGTGAACTGG 790
Db 8365 TGTCTGCGCCACGCGCGGCAAGACACACCTCGAGCGGGGGCGCGGACGACATCTGTGTCG 8424
Oy 791 CCGACACAGATGGGATTCGAAAGCCACCCCTGCGGGAGCTCACACAGCCGTGAATCTACAGAG 850
Db 8425 GCGCGCGCGGGCGGAGAGCTTCTACCGCGGGCGCGGAGCGGAGCGTGTTCGCTTCGAGC 8484
Oy 851 CCGACGAGTTGATGTGCGGTTCACCCACCGCGGGGGGTCAACACCGATCAACTCGCTGATG 910
Db 8485 CGCTGTCCGACAGCGCAGCGCACTTACGATGCGGGAGAACACAGGGGCGACCGCATCGCGC 8544
Oy 911 GCGAGCCGCTGGGCAACGGCGAGCCGGTTCACATGACGCTGGCCATCTCCGGGACCGGTTCT 970
Db 8545 ACTTCGGGGTGGGGGAGAACACCTCGAGCTATCGGCGCTGGGGCTTTCACCGGGCTGGGCG 8604
Oy 971 GGGGCGTGAATGAGACAGCGCGGGCGCGCTGATGAAACGATCGAATATG 1019
Db 8605 ACGGCTACAGCGCACCTTCGCCCTGTGTCTCAACAGCGCGCGGACCG 8653

RESULT 4
US-08-403-852D-4
: Sequence 4, Application US/08403852D
: Patent No. 5891695
:
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanche, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crey-Lagard, Valerie
: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: NUMBER OF SEQUENCES: 43
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
:
: COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/403,852D  
 FILING DATE: 10-MAY-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/FR 93/00923  
 FILING DATE: 25-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-00000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1208 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: S.pristinaespiralis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1208  
 US-08-403-852D-4

Query Match 5.6%; Score 57.4; DB 2; Length 1208;

Best Local Similarity 42.18; Pred. No. 0.0004;  
 Matches 389; Conservative 0; Mismatches 531; Indels 3; Gaps 1;

QY 16 GACCTCGGACCTTCAACCTGTGGCGTGGAGCCGCGCATCCGAGAGACCCCG 75  
 DB 73 GACACCGCTTCGACCGCCCTCTCGCGAGAGACCCCGCTCAGCGCTCGAGAC 132  
 QY 76 GCCGCTCGGTGATCCAGTACAGGAGTACAGAACTTCCAGCCGTTGCGCGG 135  
 DB 133 CTGATACACACCGGCGAGTCCACATCGCGCGAGGTACACACCAAGGCTAGCGCC 192  
 QY 136 GCGCTCGCTGATCGAGGGGATACCTCGCGCCGCAAGAGAGAGATCTCATCTTC 195  
 DB 193 ATGCCCCAATGTGTCGCGAGACAGATCTTGGCCATCGCTAGCATCTGTCGCGCAAGGGC 252  
 QY 196 GACACCGGATTCGGTCAATTCGATCTGACTACACCTCGCGCATGTATGACAGGCAAC 255  
 DB 253 TTGACGCGGCTCTCGCGGCTCTCGCTCTCATCGCGCGGCGCATGCCCGGCAATGCC 312  
 QY 256 ATCTTCGGGCTCGCGACCACTTGACCGGTTGCTCGACGGGGCTTCAAGTGGCGCTG 315  
 DB 313 CAGGGCTCGACAGCGCTACAGACCCGCGTGAAGGGGAGACGAGCTCGACAG 372  
 QY 316 GAGCGCGGTACACAGAGAGAACTG--GCCGAGATCACCAAGAGTCCGTGATG 372  
 DB 373 CAGGGCGCGCGGACGACGCGCTGATTTGGCTACCGCACGAGACCCCTCGCTG 432  
 QY 373 TCGAGCTGGCGGAATGTTCTGATCTGACGCTACCGGGGATATCGAAGAGCGCAAG 432  
 DB 433 ATGCGCGTGGCCATCGAGCTTCGCCACCGCTCTCGCGCGGCTCACCGAGGTCCGCAAG 492  
 QY 433 GCGGAGAGAGACGTGTCAAGCTCACCCATCAGGTGTATCTTACAGCCATCCGTTACTG 492  
 DB 493 GAGGAGACCGTCCCTTACGTGCGCGGAGGAGGAGACCAAGGTACACATCGATGATCAG 552  
 QY 493 TGGGCTTCCCGCGCGGAGACAGATCTTGGGACACACCGGATCGTCCGCGCATGTC 552

DB 553 GCGAGCGCCCGGTGCGCTGAGACACCGTCTGCTCTCCAGACGCCCGCGCATTC 612  
 QY 553 CGCGCGCGCGGCGCAACACCGTTCGACCGGACCATCAAGAACTACAGTGGGTGATTC 612  
 DB 613 GACTGGGCTCTGCTCTACACCGCGGACATCCCGAGACAGTGTGAGACATCTCTCGCC 672  
 QY 613 ACCGAGCCAGTTTGAAGCAAGACCGTGTGCGCGGACCGGATCTGCTGCACTG 672  
 DB 673 GCACTGCGCGAGGAGCATCAAGTCTGAGAGAGACGACATACCGCTGTGTCAACCGG 732  
 QY 673 GACAACTGCGTGGCCCAAGTTCGCGGCTTCAAGTGTGATGTCAAGAGCGCAACTG 732  
 DB 733 ACCGCGCTTTTGAATTCGCGGCGCGATGAGGAGACCGGCTGACCGCAAGATC 792  
 QY 733 GCTCCCGCTCCGGAACCGTTCGCGGCGCATTCACCGGATGATGTTGCACTGAGCC 792  
 DB 793 ATCATGACACGTAACGCGCGGATAGCGCCCGACAGGGGTGCGGCTTCTCCGCAAGAC 852  
 QY 793 GACCAATGGGCAATGCAAGCCACTGCGCGACGTCACCGCGGATGATGATGATGATG 852  
 DB 853 CCGTCCAGGTGACCGCTTCCGCGGATGATGATGATGATGATGATGATGATGATG 912  
 QY 853 GACGATTTGATGGCGGTACACACCGGCGGCGGTACACCATCACTGCTGATGATG 912  
 DB 913 GCGCGGCGCTGCGCTTCCGCGGAGTTCAGGTGCTTACCGCATGCGCAAGCGCGAG 972  
 QY 913 GAGCGCGTGGGCAACGCGGAGCC 935  
 DB 973 CCGGTGCGCTTCTCTCGAGAC 995

# RESULT 5

US-08-510-646B-4  
 ; Sequence 4, Application us/08510646B  
 ; Patent No. 6077699

## GENERAL INFORMATION:

APPLICANT: Blanc, Veronique  
 APPLICANT: Blanc, Francois  
 APPLICANT: Crouzet, Joel  
 APPLICANT: Jacques, Nathalie  
 APPLICANT: Lacroix, Patricia  
 APPLICANT: Thibaut, Denis  
 APPLICANT: Zagorec, Monique  
 APPLICANT: Debussche, Laurent  
 TITLE OF INVENTION: De Crecy-Lagard, Valerie  
 TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flinagan, Henderson, Farbow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.

COUNTRY: USA

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/510,646B

FILING DATE: 03-AUG-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/403,852

FILING DATE: 10-MAY-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR 93/00923

FILING DATE: 25-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-01000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1208 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOPHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: S.pristinaespiralis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1208  
 US-08-510-646B-4

Query Match 5.6%; Score 57.4; DB 3; Length 1208;  
 Best Local Similarity 42.1%; Pred. No. 0.0004;  
 Matches 389; Conservative 0; Mismatches 531; Indels 3; Gaps 1;

16 GACCTGGGACCTTCACACCTGCTGGCCGTCGAGCCCGGCGCCATCCCGGAGACACCCCG 75  
 73 GACACCTCTCTCCAGCCCTGCTGCGGAGACCCCGCTCAGCCGCTGCGGTCGAGACC 132  
 76 GCGGCTCGGTGATCCAGTACGACGACTAGCACTGAGACCTCCAGCCGCTGCGCGGC 135  
 133 CTGATCACACCGCCGAGCTCCACATCGCCGGGAGGTACACCAAGGCTGACGCGCC 192  
 136 GGGCTGCTGATTCGAGGCGGAATACCTGCGCGGAGAGAGCAAGATCTTCATCTTC 195  
 193 ATGCGCCACTGCTGCGCGACAGATCTGCGCATCGGCTACGACATCTCCGCGCAAGGG 252  
 196 GACACCGGATTCGCTCATTCCTGATCTGACCTACACCTGCGGCTATATGCGACGGC 255  
 253 TTGAGAGCGCCCTCTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 312  
 256 ATCTTCGCGCTCGCGACACCTGACCGGCTGCTCGACGGGCGCTCCAGCTGCGCTG 315  
 313 CAGGCGCTGACAGCGCTTACAGAGCCGCTGAGGGGAGAGACACGATCTCGACAG 372  
 316 GACGCGGGTACAGCAAGAGCACTG--GCCGAGATCACCAAGAGTGGTGTGATG 372  
 373 CAGGCGCGCGGACGAGGCGCTGATGTTGCTACGCGCACGAGAGACCCCTGCTG 432  
 373 TCCGAGCTGGCGAATGTTGATGATCTGACCTGACCCGCGGAGATACGGAAGCGCAG 432  
 433 ATGCGCTGCTGATGAGCTGACCGCCACCGCTCTCGCGCTGCTGACCGAGTCTCGCAG 492  
 433 GGGAGAGAGACCTGTTCACAGCTACCCATCAGGTATCTACGCAATCCCTGACTG 492  
 493 GAGGCGACCGTCCCTTACCTGCGCGCGCGGAGGAGACCGAGTCAACATCGATGACAG 552  
 493 TGGGCTTCCCGCGCGCGGAGAGATTTGCGACACACCGCGATCGTGGCGCGCATGTC 552  
 553 GCGAGCGCGCGCGGAGCGCTGAGACCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 612  
 553 CGCGCGCGCGCGGAGACCGCTGAGACCGCATCAAGAACTACAGAGTGGGTGATCTC 612  
 613 GACCTGGGCTCTGCTCTACCGCCGAGATCTCGGAGCACTGCTCGACGCTCTCTGCT 672  
 613 ACCGACACCAATTCGAGGCAAGAGACCTGATGTCGCGGACCGCGATCTCTGCTGACTG 672  
 673 GCACTGCGCGGAGAGCGCATCAAGCTGAGAGGAGACAACTACCGCTGCTGCTCAACCGG 732

673 GACAACTGCTGCGCGGAGAGTCCGGGCTTCAACGTGTCATCTGCAAGAGCGCAAGCTG 732  
 733 ACCGCGCTCTTTCAGATCGCGGCGCCGATGAGCGACGCGGCTGACCGCGGCAAGATC 792  
 733 GCGTCCCGGTCGCGGAGCGGTTGCGCGGCAACACCCCTAAGAGCGGTGTGCAACTGACC 792  
 793 ATCATTGACACGCTACGCGGCGATGCGCCGACAGCGGCTGCGGCGATCTCCGCGCAAG 852  
 793 GACGAGATGGGATGGAAGCAACCTGCGCGACGTCACACCGCTGAACTCTAGACAGCGC 852  
 853 CCGTCAAGGTGACGCTTCCGCGCGCTAGAGGATGCGCTGGTCCGCAAGAGCGTCTC 912  
 853 GACGAGTGTGCGCGCTACCAACCGCGCGGCGGCTCAACCGATCACTCGCTGATGCG 912  
 913 GCGCGGCGCTGCTCCCGCTGCGAGGTCAGGTGCGCTACGCGCATGCGGAGCGCGAG 972  
 913 GAGCGCTGGCGCAAGGCGAGCC 935  
 973 CCGTGGCGCTTCTGTCGAGAC 995

RESULT 6  
 US-09-231-818-4  
 Sequence 4, Application US/09231818  
 Patent No. 6171846  
 GENERAL INFORMATION:  
 APPLICANT: Blanc, Veronique  
 APPLICANT: Blanche, Francis  
 APPLICANT: Crouzet, Joel  
 APPLICANT: Jacques, Nathalie  
 APPLICANT: Lacroix, Patricia  
 APPLICANT: Thibaut, Denis  
 APPLICANT: Zagorec, Monique  
 APPLICANT: Debussche, Laurent  
 APPLICANT: De Crecy-Lagard, Valerie  
 TITLE OF INVENTION: Polypeptides Involved in The  
 TITLE OF INVENTION: Biosynthesis of Streptogramins, Nucleotide Sequences  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Flanagan, Henderson, Farbow, Garrett & Dunner  
 STREET: 1300 I Street, N.W., Suite 700  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZIP: 20005-3315  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/231,818  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/403,852  
 FILING DATE: 10-MAY-1995  
 APPLICATION NUMBER: PCT/FR 93/00923  
 FILING DATE: 25-SEP-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FR 92/11441  
 FILING DATE: 25-SEP-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Meyers, Kenneth J.  
 REGISTRATION NUMBER: 25,146  
 REFERENCE/DOCKET NUMBER: 03806.0054-00000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 408-4000  
 TELEFAX: (202) 408-4400  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1208 base pairs  
 TYPE: nucleic acid





Db 3630 GACACGCTCCTCCAGCCCTGCTGCGGAGAGACCCCGCTCAGCGCTCCGCTGAGACC 3689  
 QY 76 GCCGCTCGGTGATCCAGTACAGGACTACGAATGACACCTCCAGCCGCTTCCGCGGC 135  
 Db 3690 CTGATGACACACCGGCGAGGTCCACATGCGCGGAGGTCCACCAAGGCGTACGCGCC 3749  
 QY 136 GCGCTCCCTGATCGAGGGCGAATACCTCCGCGCGAAGAACGGAATCTCTCATCTTC 195  
 Db 3750 ATCCGCACTGCTGCTCCGACAGATGCTCCGATCGGCTACGACTGCTCCGCAAGGGC 3809  
 QY 196 GACACCGGATTCGCTCATTCGATCTGATACACGCTCCGATGATGATGACAGGCAAC 255  
 Db 3810 TTCCACGCGCGCTCTCCGCGCTCTCCGCTCTCCATGCGCGCGCAATGCCGACATCGCC 3869  
 QY 256 ATCTTCGCGCTCGGCGACCACTGAGACCGGTTGCTCGACGGCGCTCCAGTCCGCGCTG 315  
 Db 3870 CAGGCGCTCGACAGCGCTTACAGAGACCGGCTGAGAGGCGAGAGACGAGCTCGACAG 3929  
 QY 316 GACGCGCGGTACAGCAAGAGAGAACTG--GCCGATATCCAGAAAGTCCGCTGATG 372  
 Db 3930 CAGGCGCGCGCGCGACGCTGATGCTGAGCTTACGCGCGACGAGACGACCCCTGCTG 3989  
 QY 373 TCCGAGCTGCGGCAATGCTTCTGTAATCTGACGCTACCGCGGATACGGAAGCGCAAG 432  
 Db 3990 ATGCGCTGCGCCATCGAGCTCGCCACCGCTCTCCGCGCGCTCACCGAGTCCGCGAAG 4049  
 QY 433 GCGCAGAGAGACCTGCTCAAGCTCACCCATCAGGTGATATCTACGCGCATCCGTAAGT 492  
 Db 4050 GAGCGCAACCTGCTTCTGCTGCGCGCGCGAGCGCAAGCCAGGTTCAGCATCGATGACAG 4109  
 QY 493 TGGGCTTCCGCGCGCGCGCGAGAGATCTTGGGCAACCGCGCATCGTCCGCGCGCATGTC 552  
 Db 4110 GCGACGCGCGCGCGCTGCGCTGAGACCGGTGCTGCTCTCCAGACGCGCGCGCATTC 4169  
 QY 553 CGCGCGCGCGCGCGCAACCGCTGACCGCGCGCATCAAGAACTACAGTGGGGTGAATCTC 612  
 Db 4170 GACTCGCGCTCCCTGCTGCTACCGCGCGCATCCGCGAGACCTCTCGAGAGCACTGCTCGCC 4229  
 QY 613 ACCGAGCCGATTCGAGGCGCAAGAGACCGGTGCGCGCACCGCGCATCTGCTCGACTCG 672  
 Db 4230 GCACTCCCGAGAGCGGATCAAGCTGAGACGAGCAACTACCGCTGCTGCTCAACCCG 4289  
 QY 673 GACAACTGCTGCGCGCAAGTCCGCGCTTCAAGTGTGCACTGCTCAAGGAGCGCAAGCTG 732  
 Db 4290 ACCGCGCGCTTCCGATCGCGCGCGCGATGAGGCGCGCGCGCTGACCGCGCGCAAGATC 4349  
 QY 733 GCTTCCCGCTCCGAGAGCGGTTGCGCGGATCACCCGTAAGAGGTTGTAAGTGGCC 792  
 Db 4350 ATCATCGACAGCTACGCGCGCGATGCGCGCGAGCGGCGGTGCTTCCGCGCAAGGAC 4409  
 QY 793 GACCAATGCGGATCGAGCGCAACCTTCGCGAGCTCACAGCGGTGATCTTACGAGCGC 852  
 Db 4410 CCGTCCAGAGGTCGACCTTCCGCGCGCTACGCGATGCGGTGCGCGCAAGAGAGTCTGTC 4469  
 QY 853 GAGCAGTTGATGCGGCGACACCGCGGCGGCGGTGACACGATCACTGCTGAGTGGC 912  
 Db 4470 GCGCGCGCGCTGCGCTCCGCTGCGAGGTGCGCTTACGCGCATGCGCAAGGCGGAG 4529  
 QY 913 GAGCGCGTGGGCAACGCGCGAGCC 935  
 Db 4530 CCGGTGCGCGCTGCTGCTGAGAC 4552

RESULT 8  
 US-08-510-646B-1  
 ; Sequence 1, Application us/08510646B  
 ; Patent No. 6077699

; GENERAL INFORMATION:  
 ; APPLICANT: Blanc, Veronique  
 ; APPLICANT: Blanche, Francis  
 ; APPLICANT: Crouzet, Joel  
 ; APPLICANT: Jacques, Nathalie  
 ; APPLICANT: Lacroix, Patricia

; APPLICANT: Thibaut, Denis  
 ; APPLICANT: Zagorec, Monique  
 ; APPLICANT: Debussche, Laurent  
 ; APPLICANT: De Crecy-Lagard, Valerie  
 ; TITLE OF INVENTION: Polypeptides Involved In The  
 ; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences  
 ; TITLE OF INVENTION: Coding For These Polypeptides And Their Use  
 ; NUMBER OF SEQUENCES: 45  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flanagan, Henderson, Farbow, Garrett & Dunner  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/510,646B  
 ; FILING DATE: 03-AUG-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/403,852  
 ; FILING DATE: 10-MAY-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/FR 93/00923  
 ; FILING DATE: 25-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: FR 92/11441  
 ; FILING DATE: 25-SEP-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Meyers, Kenneth J.  
 ; REGISTRATION NUMBER: 25,146  
 ; REFERENCE/DOCKET NUMBER: 03806,0054-01000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5392 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: S.pristinaespiralis  
 ; US-08-510-646B-1

Query Match 5.6%; Score 57.4; DB 3; Length 5392;  
 Best Local Similarity 42.1%; Pred. No. 0.00053;  
 Matches 389; Conservative 0; Mismatches 531; Indels 3; Gaps 1;

QY 16 GACCTCGGACCTCCACCTGCTGCGCTGAGCGCGCGCGCATCCGAGAGACACCCCG 75  
 Db 3630 GACACGCTCCTCGACGCTCGCTGCGCGAGAGACCCGCTTCAAGGTGCTCGAGAGCC 3689  
 QY 76 GCGGCTCGGTGATCCAGTACAGCGGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 135  
 Db 3690 CTGATCACCACCGCGCGGCTCCACATGCGCGGAGAGTCCACCAAGGCGTACGCGGCC 3749  
 QY 136 GCGTTCGCTGATCGAGGCGCAATACCTGCGCGCGCGAGAGAGAGAGAGAGATCTCTTC 195  
 Db 3750 ATGCGCCCACTGCTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3809  
 QY 196 GACACCGGATTCGATTCCTGATCTGACCTTACACCGCTGCGCGCATGTATGACAGGCAAC 255  
 Db 3810 TTCCAGCGCGCGCTGCTGCGCGCTCTCCGCTCATTCAGCGCGCGAGTCCCGGAGATCGCC 3869

```

OY 256 ATCTCCGGCTGGGACACCTGACCGGTTCTCGACGGGGCTCCAACTGGCCTG 315
Db 3870 CAGGGCGTGACAGCGCTAGAGACCGCGTCGAGGGAGAGACGACGCTGACGACG 3929
OY 316 GACGGCGGGTACAGCAAGAGCAACTG---GCCAGATCACCAGAAAGTGGCTGCATG 372
Db 3930 CAGGGCGGGGAGACAGAGGCTGATGTTGCGCTACGCGACGAGAGACCCCTGCTG 3989
OY 373 TCGCAGCTGGCGGAATCTGTAATCTGACCGCTGACCGGGGATACGAAAGCGCAAG 432
Db 3990 ATGCCGCTGCGCATGACGACGCGCCACGCGCTCTGCGCGCGGCTACAGAGTCCGCAAG 4049
OY 433 GCGGGAAGAGACCTTCCAACTGACCATCAGGATGATCATCTAGGCTATCCCTACTG 492
Db 4050 GACGGGACCGTCCCTTACTCTGCGCCGACGAGACCGAGTCCACTGATGATCCAG 4109
OY 493 TGGGCGCTTCCGCGCGCGGACAGATCTTGGGACACCGGATGCTCCCGGCGATGTC 552
Db 4110 GGCAGCGCGCGCGGTCGCTGACACCGCTCTGCTCTCTCCAGACGACGCGCGGACATC 4169
OY 553 CGCGCGCGCGCGCGGACACCGCTGACCGGACCATCAGAACTACAGTGGGATGATCTC 612
Db 4170 GACCTCGGCTCCCTGCTGACCGCGGACATCGCGGACGATGCTGACGACGCTGCTG 4229
OY 613 ACCGAGCAGATTTGAAAGCCAGAGACCGTGTGCGCGGACCGGATCTGCTGACTGC 672
Db 4230 GCACCTCGCGGAGAGCGGATGAAGTCCAGAGGACAACTCCCTGCTGCTGACACCG 4289
OY 673 GACAACCTGCTGGCGGAGAGTCCGGGCTTCAACGCTGCTGCTGCTGCTGCTGCTG 732
Db 4290 ACCGCGCGCTTTCGAGATGGCGGCGCGGATGGGCGGACCGCGGCTGACCGCGGAAATC 4349
OY 733 GCTTCCCGCTCCCGGACACCGCTTGGCGGACATCACCCTGAAGCGTGTGCTGAACTGC 792
Db 4350 ATCATCGACAGTACGCGGCGGATGCGCGGCGGCGGCGGCTGCTGCGGCAAGAGAC 4409
OY 793 GACGAGATGGGAGATGAGACCGCTGCGGACGCTGACACACCGCTGAACCTTAGAGAGCC 852
Db 4410 CCGTCCAAAGTGCACCGCTTCCGCGCGGATGCGGATGCGGCTGCTGCGGCAAGAGTGTG 4469
OY 853 GACGAGTGTGATGGGCTGACACCGCGGCGGCTGACACCGGATCACTGCTGATGTC 912
Db 4470 GCGGCGGCGCTGCGCTGCGGCTGCGGAGTCCAGGCTGCTGCGGATGCGGAGAGGCGGAG 4529
OY 913 GAGCGCTGGGCAAGCGGAGCC 935
Db 4530 CCGGCTGGCGCTGCTGCTGCGGAGAC 4552

```

## RESULT 9

US-09-231-818-1

Sequence 1, Application US/09231818

Patent No. 6171846

## GENERAL INFORMATION:

```

APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crozet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crey-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Stereogramins, Nucleotide Sequences
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA

```

```

ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,818
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
US-09-231-818-1
Query Match 5.6%; Score 57.4; DB 4; Length 5392;
Best Local Similarity 42.1%; Pred. No. 0.00053;
Matches 389; Conservative 0; Mismatches 531; Indels 3; Gaps 1;
OY 16 GACCTGGGACCTCAACCTGCTGGCGCTGACGCGCGGCGGCGGCGGCGGCGGCGGCGG 75
Db 3630 GACACCGCTCTGACCGCTGCTGCTGCGGAGGACCGCGCTGACGCGCTGACGAGAC 3689
OY 76 GCGGCTCGGTGATCCAGTACAGCGACTACGAACTGACACCTCCAGCGCTGCGCGG 135
Db 3690 CTGATACACACCGCGGCGGATCCAGATCGCGGCGGAGGTACACCAAGCGTACGCGCC 3749
OY 136 GCGGCTCGGTGATCCAGTACAGCGACTACGAACTGACACCTCCAGCGCTGCGCGG 195
Db 3750 ATGCGCCAACTGGTCCGACACAGATCTGCGCATGCGGTACGACTGCTGCGCAAGGGC 3809
OY 196 GACACCGGATTTGGTATTCGATCTGACTTACACCGTCCGCGCATGATGTCAGCGGCAAC 255
Db 3810 TTGAGAGGGGCTCTCGCGGCTCTCCATCTGCGGCGGCGGCGGCGGCGGCGGCGG 3869
OY 256 ATCTCCGGCTGGGACACCGTGAACCGGCTGCTGACGGGGGCTCCAAAGTGGCGCTG 315
Db 3870 CAGGGCGTGACAGCGCTAGAGACCGCGTCGAGGGGAGAGACGAGTCCGACGACG 3929
OY 316 GACGGCGGGTACAGCAAGAGCAACTG---GCCAGATCACCAGAAAGTGGCTGCATG 372
Db 3930 CAGGGCGGGGAGACAGAGGCTGATGTTGCGCTACGCGACGAGAGACCCCTGCTG 3989
OY 373 TCGCAGCTGGCGGAATCTGTAATCTGACCGCTGACCGGGGATACGAAAGCGCAAG 432
Db 3990 ATGCCGCTGCGCATGACGACGCGCCACGCGCTCTGCGCGCGGCTACAGAGTCCGCAAG 4049
OY 433 GCGGGAAGAGACCTTCCAACTGACCATCAGGATGATCATCTAGGCTATCCCTACTG 492
Db 4050 GACGGGACCGTCCCTTACTCTGCGCCGACGAGACCGAGTCCACTGATGATCCAG 4109

```

QY 493 TGGGCTTCCCGCCGCGAGACAGATCTGGGACACACCGGAGATGCGCGGCATGTC 552  
 Db 4110 GCGAGCGCCGCGGCGCTTGAGACACCGTGTCTGCTCCGACGACGCGCGACATC 4169  
 QY 553 CGCGCGCGCGCGCGACACCGGTGACCGACATCAAGAACTACAGTGGGGTGTATCTC 612  
 Db 4170 GACCTCGGCTCCCTGCTACACCCCGACATCCGCGAGCAGCTGTCGACACGTCCTGCC 4229  
 QY 613 ACCGAGCAGATTTCGAAGACGAGACGCTGTCGCGACCGGACGATCTGCTGACGTCG 672  
 Db 4230 GCACTCGCGAGAGACGCGATCACTGAGACGACAACTACCGCTGCTGTATACCCG 4289  
 QY 673 GACAACCTGCTGCGGAGATCCGCGCTTCAACGTGTGATCTGAAGAGACGCGACGTCG 732  
 Db 4290 ACCGCGCGCTTTCGAGATCGCGCGCGCGATGAGCGACGCGCGCGCGCGCGACATC 4349  
 QY 733 GCTTCCCGCTCCGAGACGCTTCCGCGGATCACTACCGGTAAAGAGGTTCGAACTGCGC 792  
 Db 4350 ATCATCGACACGTACGCGCGGATGCGCGCGCGCGCGCGGCTTCTCGCGAAGGAC 4409  
 QY 793 GACGAGATGGGATCGAAGACCGCTGCGGAGCTGACACGACGCTGAATCTACGACGCG 852  
 Db 4410 CCGTCCAGAGTTCGACCGCTTCCGCGCGCTGACGACGCTGCGTGGCGCGAAGAGTGTG 4469  
 QY 853 GACGAGTGTATGCGGCTCACACCGCGCGGCGGTCAACCGATCAACTCGCTGATGCG 912  
 Db 4470 GCGCGGCGCTCGCGCTCCGCTGCGAGGTCCAGTCCGCTACGCGCATCGCGAAGCGCGAG 4529  
 QY 913 GAGCGCTGGGCGACGCGGCGCG 935  
 Db 4530 CCGTCCGCGCTGCTGTCGAGAC 4552

## RESULT 10

US-09-023-173-4  
 ; Sequence 4, Application US/09023173  
 ; Patent No. 6066781

## GENERAL INFORMATION:

APPLICANT: Sulliff, Thomas D.  
 APPLICANT: Rodriguez, Raymond L.  
 TITLE OF INVENTION: Production of Mature Proteins  
 TITLE OF INVENTION: In Plants  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dehlinger & Associates  
 STREET: 350 Cambridge Ave., Suite 250  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/023,173

FILING DATE: 13-FEB-1998

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/038,168

FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: Pelichory, Joanne R

REGISTRATION NUMBER: P42995

REFERENCE/DOCKET NUMBER: 0665-0007.30

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-0880

TELEFAX: 650-324-0960

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1140 base pairs

TYPE: nucleic acid

STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: codon-optimized Ram3d signal-prosult111slin BPN  
 US-09-023-173-4

Query Match 5.6%; Score 57; DB 3; Length 1140;  
 Best Local Similarity 46.3%; Pred. No. 0.00049;  
 Matches 232; Conservative 0; Mismatches 260; Indels 9; Gaps 1;

QY 447 GTCCAGCTCACCCATCAGTGTATCATCTACGATCCCGGTACCTTGGCGCTTCCCGCC 506  
 Db 468 GACCAACCGCTTTCAGACACCAACAGCATGACACCGTCCGCGACCGCTGCGCGC 527  
 QY 507 CGCGAGCAGATCTTGGGACACACCGCATGCTGCGCGCATGCTGCGCGCGCGCGCGCG 566  
 Db 528 CTCACCAACAGATGAGGGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 587  
 QY 567 CAACACCGCTGACCGCGACATCAAGAACTACAGTGGGGTGTATCTACCGCGACGCTT 626  
 Db 588 GGTACTCGGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 647  
 QY 627 CGAAGCGCAAGACCGTGGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686  
 Db 648 CATCGCACCAACATGAGTACATCATCATGACCTTCCGCGCGCGCGCGCGCGCGCGCG 707  
 QY 687 CGAAGGTCGCGCGCTTCAACGTGTGATCTGATCAAGAGCGCGCGCGCGCGCGCGCG 746  
 Db 708 CTCACAGGCG 767  
 QY 747 GAACGCGTTCG 806  
 Db 768 CAACGAGGCG 827  
 QY 807 CGAAGCG 857  
 Db 828 CATCG 887  
 QY 858 GTTGTATGCG 917  
 Db 888 GAGCTGAGCTATGCG 947  
 QY 918 CGTGGCGCACG 938  
 Db 948 CCGCGCGCTTACG 968

## RESULT 11

US-09-173-300-21

; Sequence 21, Application US/09173300

; Patent No. 6451581

GENERAL INFORMATION:

APPLICANT: Falco, Saverio Carl

APPLICANT: Hitz, William D.

APPLICANT: Kinney, Anthony J.

APPLICANT: Cahoon, Rebecca E.

APPLICANT: Ratajski, J. Antoni

TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES

FILE REFERENCE: BB-1126

CURRENT APPLICATION NUMBER: US/09/173,300

EARLIER FILING DATE: 1998-10-15

EARLIER FILING DATE: 1997 October 28

NUMBER OF SEQ ID NOS: 54

SOFTWARE: Microsoft Word Version 7.0A

SEQ ID NO 21

LENGTH: 1162

TYPE: DNA

ORGANISM: Glycine max

US-09-173-300-21

Query Match 5.4%; Score 55.4; DB 4; Length 1162;  
 Best Local Similarity 49.9%; Pred. No. 0.0011;

; STRANDEDNESS: single

QY	35	TGTGGCCGTGTGACCCGGGCGCCTTCGGCAGAGACACCCGGGCGGCTCGTATCCAGT	94
Db	196	TGTATCGACCTTACCCCTTCGCGACCCCGGGCGGCGCCCTCCAGGCCCGGCTCCACC	255
QY	95	ACAAGCACTACGAACTGGACACTCCAGCCGTTTCGGCGGGCGGCTGACCTGGATCGAG	154
Db	256	GCAGGGGCGCGGTACCGCGCGCCCTTCACCGCGGCGGCTCGTGGCGGCTCGGGCTCGCC	315
QY	155	GGCAATACCTGCGCGCGGCGGAGAGCAAGATCTGCATCTTCGACACCGATTGGTCAATT	214
Db	316	TCACCGGCGCGCGACCGCGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	375
QY	215	CCGATCTGCACCTTACACCGCTGCGCATGTATGGACCGGACGACATCTTCGGATCGGGAC	274
Db	376	TACCGGCGCCACCGGCTGACCGGTTACTGCGAGAACTTCAACAACGGCGGACCGTCAACA	435
QY	275	ACCTGACACCGGTTGCTGCAGCGGGGCGGTCGAAGCTGCGCTTCGACGCGCGGGTACGCAAG	334
Db	436	CCCTTCGCGACGTCGCGAGCGGCTACGACATCATCGCGCGCTTCCTTCGCGCAAGCCACGG	495
QY	335	ACGAACTGGCGGAGATACCAAGAAGTGGCTGTGCATGTGCGAGCTGCGGGAATGCTTCG	394
Db	496	CCAAAGGCGGGGAGATACCTTCACCTTCGACTCGGCTCGGCGGCTGACACGACGAG	555
QY	395	TGAATCTGACCGCTGACCCCGGGGATATAGGAAAGGCGAAGGGCGAGAGACGTCGTCAAC	454
Db	556	AGCAGTTTCGGCGCGACCTTCGCGCCACGACAGCGCCGACGCAAGTGTGTATCATCTCG	615
QY	455	TCACCCATCAAGTGTATCATCTACGCCATCCCTGACTGTGTGGGCTTCGCGCCCGCGAC	514
Db	616	TCGCGCGCGAAGAGGGCGGCGGTGCGGTCAACGACAGCGGCTCGCGCCAGCGGTTGCGG	675
QY	515	AGATCTTCGGGACACCGCGCATGTGCGCGGCAATGTCCGCGCGCGCGCGCGCAACCG	574
Db	676	ACACGACCTACGGGCT--GATGAGAGAGTACGGCTTCGACGCGGCTCGACATCGACTGG	732
QY	575	TCGACCGGACCATCAAGACTATACGAGGGGATGATTCGACCGGACGACATTTGCAAGCA	634
Db	733	AGAAAGGCTCTCACTCCACTATATACGAGGCGCTTCACCAAGCTTCACGAGAAAGCGG	792
QY	635	AGGACCGTGGTGGCGGACCGCGCATCTGCTGACTCGGACAACTGCTGGCGCGAAGTTC	694
Db	793	GGGAGCGCTGTCCTCATCATCATGCGCGGAGACATGCAATGCAATGCGTGGCCGGAAGC	852
QY	695	CGGGCTTCAACTGTGTGATGTGTAAAGAGCGGCAAGCTGGCTTCGCCGTCCGGAACGGT	754
Db	853	AGTACTTCAAGAGCGGCGCTGTGTACAGAAAGTCTCTGACCGCGCTCAACATGCACTACT	912
QY	755	TGCGGGGCTCACCCGTAAGAGGCTTGGAACTGGCGGACCAAGATGGGCAATCGAAGCA	814
Db	913	ACAACAGCGGCTGTGATGCTCGCTGCGGACGCGCATGGTCTACGGCGCAGGGCACCTGCACT	972
QY	815	CCCTGCGGACGTACCAAGCGCGTAACCTGTAGACGCGCACAGCAAGTGTGATGGCGGTCA	874
Db	973	TCTGCAACCGGCGTGGCTGTGATCCACTTGGAGAAGCGTGTGCAACCGCTCCCAAGTGGCA	1033
QY	875	CCGCGGGCGGGTTCACACCGATCACTCGCTGGATGGCGAGGCGCGTGGGCAACGGCGAGC	934
Db	1033	TCGGTGTCCCGCTCCCGGAAGCGGCGGGGCGGCGTACGTGAGCCCTCCGTTGTC	1092
QY	935	CCGGTCCACTGAGGCTGGCCATCCGGGAC	963
Db	1093	ACGACGCGGTGACTGCTGACCCGGGGC	1121

RESULT 13  
US-08-196-218-31  
; Sequence 31, Application US/08196218  
; Patent No. 5614619  
; GENERAL INFORMATION:  
; APPLICANT: Piepersberg, Wolfgang  
; APPLICANT: Stockmann, Michael  
; APPLICANT: Taleghani, Kamalz Mansouri  
; APPLICANT: Distler, Jurgen  
; APPLICANT: Grabley, Susanne  
; APPLICANT: Sichel, Petra  
; APPLICANT: Brau, Barbara  
; TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes  
; TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; ADDRESS: Dunner  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: United States  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/196,218  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ogden, Stasia L.  
; REGISTRATION NUMBER: 36,228  
; REFERENCE/DOCKET NUMBER: 02481.1372-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 31:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2634 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 3..401  
; NAME/KEY: CDS  
; LOCATION: 416..1531  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1561..2625  
; US-08-196-218-31  
Query Match 5.3%; Score 53.6; DB 1; Length 2634;  
Best Local Similarity 43.2%; Pred. No. 0.0031;  
Matches 430; Conservative 0; Mismatches 549; Indels 17; Gaps 3;

198 CACCGGATTCGATCCATTCGATCTACCTACCGCTCGCGCATGTATGACGACGCAACAT 257  
1821 CACCGGATTCGATCCATTCGATCTACCTACCGCTCGCGCATGTATGACGACGCAACAT 1880  
258 CTTCCGCGCTCGGCGACACCTGACCGGTGCTGACGAGGGGGGTCCAGCTGCGCTGGA 317  
1881 CACATATCGTCCCGACGAGGATCAGGATTCGAGAGGATTCACCCGCGACGCGCGCG 1940  
318 CCGCGGATTCAGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGAT 377  
1941 CCGCGGATTCAGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGAT 2000  
378 GCTGCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 437  
2001 CCGCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 2060  
438 GAAAGACCTGTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCG 497  
2061 CCGGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 2120  
498 CTTCCGCGCTCGGCGACACCTGACCGGTGCTGACGAGGGGGGTCCAGCTGCGCTGGA 557  
2121 CAGCAGCGCGCGCGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTC 2180  
558 CCGCGGATTCAGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGAT 617  
2181 GACAGCTGCGCGCGCGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAGCA 2234  
618 AGCGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 677  
2235 GACAGCTGCGCGCGCGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAGCA 2294  
678 CTTCCGCGCTCGGCGACACCTGACCGGTGCTGACGAGGGGGGTCCAGCTGCGCTGGA 737  
2295 GCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2354  
738 CCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 797  
2355 GCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2414  
798 GATGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTC 848  
2415 GCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2474  
849 CCGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 908  
2475 CCGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2534  
909 TGGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 968  
2535 CCGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2592  
969 CCGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 1004  
2593 CCGCGGATTCGCAAGGACGAGGATTCAGCAAGGATTCAGCAAGGATTCAGCAAGGATTCAG 2628

RESULT 14  
US-08-681-953-31  
; Sequence 31, Application US/08681953  
; Patent No. 5710032  
; GENERAL INFORMATION:  
; APPLICANT: Piepersberg, Wolfgang  
; APPLICANT: Stockmann, Michael  
; APPLICANT: Taleghani, Kamalz Mansouri  
; APPLICANT: Distler, Jurgen  
; APPLICANT: Grabley, Susanne  
; APPLICANT: Sichel, Petra  
; APPLICANT: Brau, Barbara  
; TITLE OF INVENTION: Secondary-Metabolite Biosynthesis Genes  
; TITLE OF INVENTION: From Actinomycetes, Method of Isolating Them, and Their

```

: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flanagan, Henderson, Farabow, Garrett &
: ADDRESSEE: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: United States
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/681,953
: FILING DATE: 30-JUL-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/196,218
: FILING DATE: 25-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Ogden, Stasia L.
: REGISTRATION NUMBER: 36,228
: REFERENCE/DOCKET NUMBER: 02481.1372-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-408-4000
: TELEFAX: 202-408-4400
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2634 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3..401
: NAME/KEY: CDS
: LOCATION: 416..1531
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1561..2625
: US-08-681-953-31

```

```

Query Match 5.3%; Score 53.6; DB 1; Length 2634;
Best Local Similarity 43.2%; Pred. No. 0.0031;
Matches 430; Conservative 0; Mismatches 549; Indels 17; Gaps 3;

```

```

OY 18 CCGGACCTCCAACTGTCGCGCCGTCGAGCCCGCGCCATCCGCGAGACACCCCGGC 77
Db 1641 CATCCCAACACACCCGTCGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 1700
OY 78 CGGTCGCGGATCCAGTACGAGCAGTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 137
Db 1701 CGAGTCGCGCTCATCTGTCGCGACCGCGCCGCGAGTCGAGCGCTGTCGCGAGAGG 1760
OY 138 CGTCGCTGATGAGGCGGATGTCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 197
Db 1761 TGCCCGGTCGACGTCGTCGATCAGTACATCCCGAGAGAGAGAGAGAGAGAGAGAG 1820
OY 198 CACGCGATTCGGTCACTTCGATCTGACCTACACCGTCGCGGATGATGAGCAGGCAACAT 257
Db 1821 CACCGTCATCAGCCCGCGGCTTCCTGCGAGCAGCAGCAGCAGCAGCAGCAGCAG 1880
OY 258 CTTCGCGGTCGCGGACACCTGAGCCGTCGTCGAGCGGCGCGTCGCAAGTCGCGCCCTGA 317
Db 1881 CAACATGTCGCGCGAGGAGTCACCGAGATGCGCGAGAGATTCAACCGCGCGCGCGCG 1940
OY 318 CGCCGCGTACAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
Db 1941 CGCCGAGTCTGTCGTCGACAGAGTCGCCGACCCGCGCTCTTCGCGGCTCGCGCAACTCG 2000

```

```

OY 378 GCTGCGGAAATCTGTCGATCTGACCCGTCACCCGCGGATACGAAAGCGCAAGGCGCA 437
Db 2001 CCGGAGCGGGAGAGTGTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2060
OY 438 GAAGGACCTTCCAAAGTCAACCATCAGTGTATCATCTACATCCGCTACCTGTGGGC 497
Db 2061 CCTGATCGGGGTCTACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2120
OY 498 CTTCGCGCGCGCGGAGAGATCTTGGCAGCCAGCGGATCTGTCGCGCATCTGCGCGC 557
Db 2121 CAGAGCGCGCGGAGAGATGAGATCACCGCGCTGTCAGTGTGTGTGTGTGTGTGTGTGT 2180
OY 558 CGCGCGCGCGGAGAGACCGTCGACCCGACATCAAGAGATCAAGAGAGAGAGAGAGAGAG 617
Db 2181 GAGAGTCGCGCGCGAGCGCTTACAGACGGCTCTGTGAGAGAGAGAGAGAGAGAGAGAG 2234
OY 618 AGCCAGTTTGAAGACCAAGAGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 677
Db 2235 GGAAGTCTTGAAGTCAAGACCACTCTGTGAGCGGCTGTACCCCGCGTGTGACGAGCA 2294
OY 678 CTGCGTCCGAGAGAGTCCGCGCTTCAACGTCGATCTGTCAAGAGAGAGAGAGAGAGAG 737
Db 2295 GGTGAGACCGCGAGACAGCTGCTGTCGCGCGGCTGTGTGTGTGTGTGTGTGTGTGTGT 2354
OY 738 CCGTCCCGGAGACCGCTTCCGCGGATCACCCGTCAGAGCGTGTGTGTGTGTGTGTGTGT 797
Db 2355 CGGTCGCGGGTGTGAGAGGCGCGCGGATCATCGCGCGGAGAGAGAGAGAGAGAGAGAG 2414
OY 798 GATGGGATTCGAAGCCACCTT-----GCGGACATCACAGCGCTGTGAATCTTACGA 848
Db 2415 GATGGGCGCGGACACCTTCATCGGGGAGATTCAGCGGAGAGAGAGAGAGAGAGAGAG 2474
OY 849 CGCGGAGAGTGTATGAGCGGTCACACCGCGGCGGAGGTCACAGCATCACTGCTGTGA 908
Db 2475 CTGATTCGCGCTGTGAGAGCGCTGTCGTCACCGCGGCTGTGTGTGTGTGTGTGTGTGT 2534
OY 909 TGCGAGCGCGGTGGGAGACCGCGGAGCGGTCATCAGAGCGTGTGTGTGTGTGTGTGTGT 968
Db 2535 CGGCGCGCGCGGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2592
OY 969 CTGGCGCTGTATGAGAGAGAGCGCGCGCGCGCTGTATCGA 1004
Db 2593 GTCGAGAGACCAACCGAGTGTGCGGCGCATGGA 2628

```

```

RESULT 15
US-09-003-287-7
; Sequence 7, Application US/09003287
; PATENT INFORMATION:
; PATENT No. 6096947
; APPLICANT: Jayne, Susan
; APPLICANT: Barbour, Eric
; APPLICANT: Meyer, Terry
; TITLE OF INVENTION: METHODS FOR IMPROVING TRANSFORMATION EFFICIENCY
; FILE REFERENCE: MOPAT_mocah
; CURRENT APPLICATION NUMBER: US/09/003,287
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Myrothecium verrucaria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (01)..(732)
; US-09-003-287-7

```

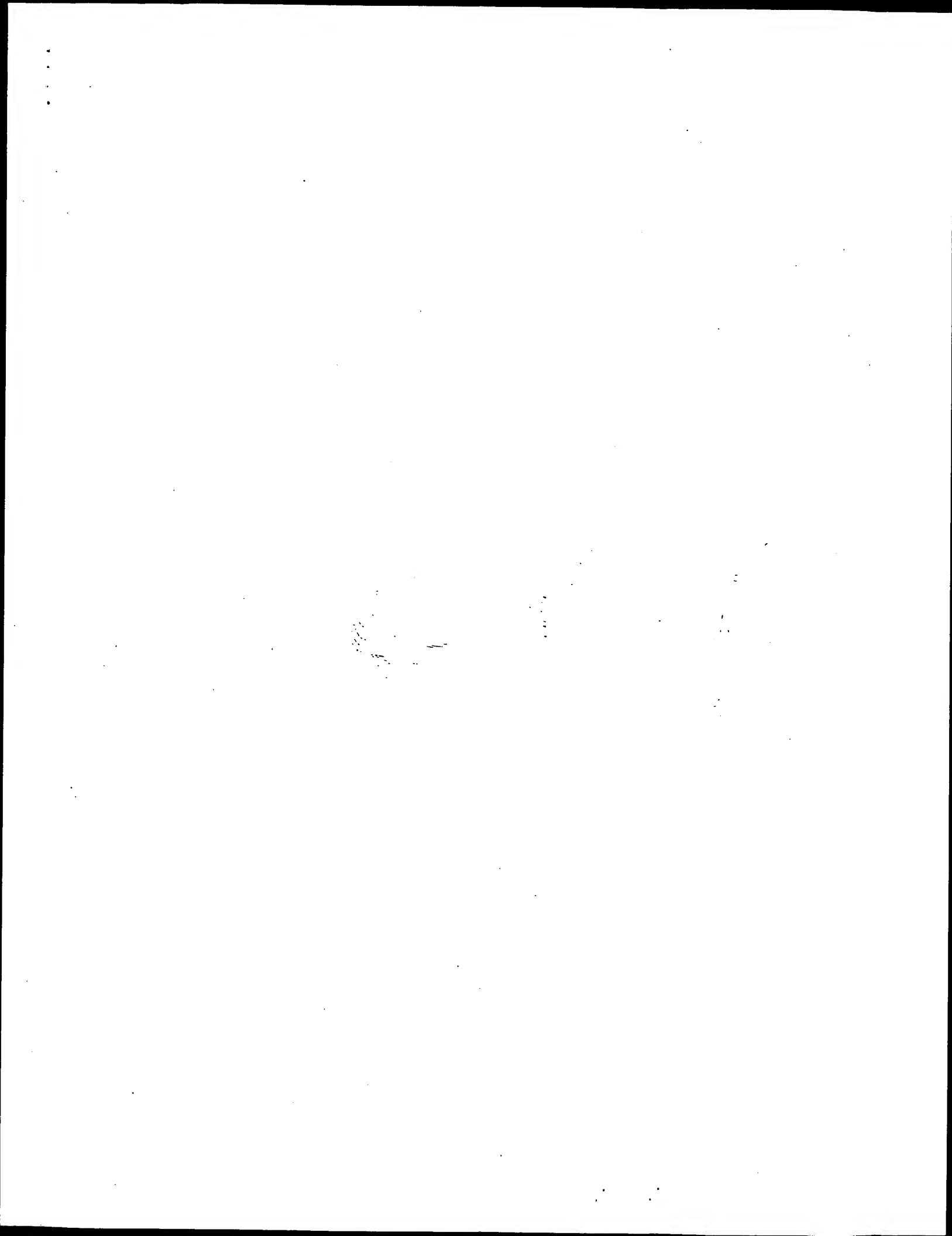
```

Query Match 5.2%; Score 53.2; DB 3; Length 735;
Best Local Similarity 43.7%; Pred. No. 0.003;
Matches 235; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

```

QY	315	GGAGCCGGGATGACGAAGAAGCAACCTGGGCGCGAGATCCACCAAGAAGGGTGTCATGTC	374
Db	57	GGCCATGTTGGACCTCCCTCGGCAAGCTCGGAGAGCTAGTCTCTACTCCGTGAGGACAT	116
QY	375	GCACGTGGCGCAATCGTTCTGTAATCTGACCGTCACCCGGGGATACGGAAAGCGCAAGG	434
Db	117	CGCCTTCCACCGGCGCAAGCTCGTGGCCGAGGCGCAAGCGCTGTGCACCCCGCCCT	176
QY	435	CGAAGAGACCTGTTCMAAGTCAACCCATCAGAGGTATCATATACGGCATCCGTACCTGTG	494
Db	177	CTCCCCGAGACCTTACACACACTCATGCGGATTTTACTTGGGGCACGCTGATCCGAG	236
QY	495	GACCTTCCCGCCGCGCGAGCAGATCTTGGGACACACCGCGCATGTGCGCGCCATGTCC	554
Db	237	GAGGTCTCTCCGAGCGAGGCGCAAGACACTCTCCCGTCACCTGGGCGCTCACCTGCT	296
QY	555	CCGGCGCGCGCGCAACACCGTCGACCGGACCATCAAGATATACCAAGTGGGGTATGTCAC	614
Db	297	CTCTCACCACGTGGGACACCGCGAGGCTTACTTACCTCCACCCGCAATGTCTCTTCGACAT	356
QY	615	CGCAGCCAGTTTCGAAGCCAAAGACCGGTGTGGCGGACCGCATCTCTCTGCATCTCGA	674
Db	357	CTACGGCGGCAATCAAGGAGTGAAGGTCTCAAGGTCTCGGGCTCTCCACCGCAAGGC	416
QY	675	CACAGTCGTGGCCGGAAGGTGCGGGCTTCAACGTGTGATGTGTCAAGGACGGCAAGCGGC	734
Db	417	GGAGCGGTTGGCGGAGGCATCATTCGCCACAGAGAGAGTGGGGCTCGACGGCAACATCAC	476
QY	735	CTTCCCGTCCCGGAAGCGGTTGCGGGCATCACCCGTAAAGACGGTTCGAACTGGCCGA	794
Db	477	CTTCTCCGCGCAGCTCATACCAAGCTTCGCCACCTCTACAGACAACTGGGCGCCTACGACGG	536
QY	795	CCAGATGGGCAATCAAGGACCACTCGGCGAGGTACACGACGCGGTGAATCTTACGACGGC	852
Db	537	CATCGACGACTTCGGGCTCTGGGTGGAGAGAACCAACCCGGAACATCATCAACACCGGC	594





GenCore version 5.1.4-p5-A578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 24, 2003, 20:08:57 ; Search time 136 Seconds  
(without alignments)  
8160.966 Million cell updates/sec

Title: US-10-067-291-2

Perfect score: 1020  
Sequence: 1 atgactgccttcagacct.....tcgaacgacgataactga 1020

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications-NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/1/pubpna/PCT05\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1020	100.0	1020	US-10-067-291-2	Sequence 2, Appl
2	325	31.9	993	US-10-112-706-6	Sequence 6, Appl
3	325	31.9	993	US-10-112-706-7	Sequence 7, Appl
4	325	31.9	1574	US-10-112-706-21	Sequence 21, Appl
5	324.6	31.8	978	US-10-112-706-5	Sequence 5, Appl
6	164.8	16.2	465	US-10-112-706-26	Sequence 26, Appl
7	66.2	6.5	15872	US-09-860-846-1	Sequence 1, Appl
8	66.2	6.5	15872	US-09-888-384B-1	Sequence 1, Appl
9	66.2	6.5	15872	US-09-861-289-1	Sequence 1, Appl
10	56.2	5.5	6798	US-09-918-740-57	Sequence 57, Appl
11	56.2	5.5	8077	US-09-918-740-63	Sequence 63, Appl
12	55.4	5.4	1162	US-10-027-450-21	Sequence 21, Appl
13	53.8	5.3	1294	US-09-748-033-2	Sequence 2, Appl
14	52.4	5.1	810	US-09-974-300-492	Sequence 99, Appl
15	51	5.0	1266	US-10-145-415-98	Sequence 14, Appl
16	50.4	4.9	1086	US-10-027-450-14	Sequence 2, Appl
17	50.2	4.9	1835	US-09-779-429-2	Sequence 5, Appl
18	49.2	4.8	905	US-09-748-033-5	Sequence 4020, Ap
19	48.6	4.8	3126	US-09-815-242-4020	

20	48.4	4.7	88421	9	US-09-976-059-1	Sequence 1, Appl
21	48.2	4.7	1473	10	US-09-070-844-25	Sequence 25, Appl
22	48.2	4.7	1506	10	US-09-070-844-23	Sequence 23, Appl
23	48.2	4.7	1969	10	US-09-070-844-7	Sequence 7, Appl
24	48.2	4.7	2096	10	US-09-070-844-19	Sequence 19, Appl
25	48.2	4.7	2099	10	US-09-070-844-3	Sequence 3, Appl
26	48.2	4.7	2137	10	US-09-070-844-18	Sequence 18, Appl
27	48.2	4.7	2140	10	US-09-070-844-1	Sequence 1, Appl
28	47	4.6	390	10	US-09-790-399-7	Sequence 7, Appl
29	46.8	4.6	822	9	US-10-101-464A-38	Sequence 38, Appl
30	46.8	4.6	1248	9	US-09-860-846-7	Sequence 7, Appl
31	46.8	4.6	1248	9	US-09-888-384B-7	Sequence 7, Appl
32	46.8	4.6	1248	10	US-09-861-289-7	Sequence 7, Appl
33	46.8	4.6	2138	9	US-10-101-464A-462	Sequence 462, App
34	46.8	4.6	12441	9	US-09-988-384B-3	Sequence 3, Appl
35	46.8	4.6	13613	9	US-09-860-846-3	Sequence 3, Appl
36	46.8	4.6	13613	10	US-09-861-289-3	Sequence 3, Appl
37	46.4	4.5	2451	10	US-09-815-242-7752	Sequence 7752, Ap
38	46.2	4.5	1503	10	US-09-815-242-7680	Sequence 7680, Ap
39	45.6	4.5	1185	10	US-09-887-576-784	Sequence 784, App
40	45.6	4.5	1609	9	US-09-931-457A-50	Sequence 50, Appl
41	45.6	4.5	8911	12	US-10-044-090-17	Sequence 17, Appl
42	45.4	4.5	1185	10	US-09-815-242-7687	Sequence 7687, Ap
43	45.4	4.5	4041	9	US-09-860-846-36	Sequence 36, Appl
44	45.4	4.5	4041	9	US-09-988-384B-36	Sequence 36, Appl
45	45.4	4.5	4041	10	US-09-861-289-36	Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-10-067-291-2  
Sequence 2, Application US/10067291  
Patent No. US20020127664A1  
GENERAL INFORMATION:  
APPLICANT: Takashima, Yoshiki  
APPLICANT: Mitsuda, Satoshi  
TITLE OF INVENTION: Protein capable of catalyzing transamination  
TITLE OF INVENTION: stereoselectively, gene encoding said protein  
FILE REFERENCE: 058251  
CURRENT APPLICATION NUMBER: US/10/067, 291  
CURRENT FILING DATE: 2002-02-07  
PRIOR APPLICATION NUMBER: 09/527, 522  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: JP 11/075511  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: JP 11/088634  
PRIOR FILING DATE: 1999-03-30  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1020  
TYPE: DNA  
ORGANISM: Mycobacterium aurum  
FEATURE:  
OTHER INFORMATION: SC-5423  
NAME/KEY: CDS  
LOCATION: (1)...(1017)  
US-10-067-291-2

Query Match 100.0%; Score 1020; DB 12; Length 1020;  
Best Local Similarity 100.0%; Pred. No. 5e-253;  
Matches 1020; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 61 CGGAGAGACACCCCGCGGCTGGTGGATCGATCAGGACTAGGACTGACACTCC 120  
OY 1 ATGACTGCTCTTTCAGACCTCGGACCTTCAACTGTTGCGGCGGCGGCAATC 60  
OY 1 ATGACTGCTCTTTCAGACCTCGGACCTTCAACTGTTGCGGCGGCGGCAATC 60  
DB 61 CGGAGAGACACCCCGCGGCTGGTGGATCGATCAGGACTAGGACTGACACTCC 120  
OY 1 ATGACTGCTCTTTCAGACCTCGGACCTTCAACTGTTGCGGCGGCGGCAATC 60  
DB 61 CGGAGAGACACCCCGCGGCTGGTGGATCGATCAGGACTAGGACTGACACTCC 120  
OY 1 ATGACTGCTCTTTCAGACCTCGGACCTTCAACTGTTGCGGCGGCGGCAATC 60

```

? APPLICANT: HASEGAWA, Junzo
? TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
? FILE REFERENCE: 1422-398P
? CURRENT APPLICATION NUMBER: US/10/112,706
? CURRENT FILING DATE: 2002-04-02
? PRIOR APPLICATION NUMBER: 09/403,493
? PRIOR FILING DATE: 1999-10-22
? PRIOR APPLICATION NUMBER: JP 9-121732
? PRIOR FILING DATE: 1997-04-23
? NUMBER OF SEQ ID NOS: 26
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 6
? LENGTH: 993
? TYPE: DNA
? ORGANISM: Arthrobacter sp.
US-10-112-706-6

```

	Query Match	31.9%	Score 325;	DB 9;	Length 993;	
	Best Local Similarity	58.2%;	Pred. No. 1.7e-74;			
	Matches 571;	Conservative 0;	Mismatches 410;	Indels 0;	Gaps 0;	
OY	40 GCCGTGACGCCGGCGCCATCCGCAGAGACAACCCTCGCGTTCGTGATCACTAACAAC	99				
Dd	13 GCCCATTAACCTCGAGATGCTTCAACGACACGACACGGCCTCGACTACATCACTTTAAC	72				
OY	100 GACTACGAACCTGANAACCTCCAGCCCCCTTGCCCGCGCTCGCTGATGAGGGCGAA	159				
Dd	73 GACTACGAACCTGATCTCTGTAAACCCGCTCCGGAGAGTGGCGCTATGATGAGGGTGCA	132				
OY	160 TACCTGCGCGCGCGAAGAAGCAAGATCTCCAATTCTCGACACCGGATGGGTCAATTCCAT	219				
Dd	133 TTTCGCGCGCGCTGGAGAGCGCGGATCTCGATCTTCGATCAAGGTTTACCTCACTCGAC	192				
OY	220 CTGACCTCAACCCCTTCGCGCATGTATGSCAGGCAACATCTTCCGGCTCGCGCACCTG	279				
Dd	193 GTCACTTCAACAGGCTTCTCCAGCTGTGAAACGAAATGATTTCCGCTGAGACCAACATC	252				
OY	280 GACCGGTGCTCGACGGGGCGTCMAAGTGGCGCTCGACGCCGGGTACAGCAAGAGCAA	339				
Dd	253 GAAGGCTCTTCTCCAAACGCGGAGTCGATGGCATCATCTCTCGCTACACAGAGCAA	312				
OY	340 CTGCGCCAGATCACCAAGAAAGTGGTGTGATGTGCGAGCTGCGCAATCTTGAT	399				
Dd	313 GTGAAGAGAGATTGGCTTCGAACTCTGCGCAAGCAAGCAATTGGTGAAGCCTTGTGTCC	372				
OY	400 CTGACCGTCAACCCGGGGATACGGAACGCGCAAGGCGGAAGAAGCTGTCCAAGTCAAC	459				
Dd	373 GTGTGATTAACCGCGGTGTTCACGCTCACTCCGGCGACGCGCATACAGAAACACCGC	432				
OY	460 CATCAAGTGTACATCTACAGCATCCCTTACCTGTGGGCTTCCGCGCGCGACACATC	519				
Dd	433 CCGAGGTGTACATGTATGGCGTCCCATATCAAGTAGATCTGTCCGTTTGACCCGATTCG	492				
OY	520 TTGGGCAACCAACCGGATCTGTGCGCGCCATGTCCGCGCGCGCGCAACACGCTGAC	579				
Dd	493 GAGCGGCTGCACGCCCAATGTGCGACAAAGGCTGCGCGCAACCCGCGACGCTGATGAC	552				
OY	580 CCAGACCATCAAGAATCACTCAAGTGGGGGATCTCAACCCGACACCACTTTCGAAGCCAGGAC	639				
Dd	553 CTCACAGTCAAGAACTCTCCAGTGGGGGATCTATCTCGTCCGGTTCAAGGAGACGACAC	612				
OY	640 CGTGTGCGCGCACCGCATCTGCTCGATCTCGGACAAATCGCTGCGCGCAAGGCTCGGCG	699				
Dd	613 CCGGGGTGAGAGCTCCCTTGTGTCTGACAGGGGATGAGATGCTTTGCGGAGGCTCGGG	672				
OY	700 TTCAAGTGTGATCTGTCAAGAGACGCAAGCTGCGCTCCCGTCCGGAACGGTTGCG	759				
Dd	673 TTCACGTGTCTGTATCAAGGACGCGCTGTGCGACGCGCGGTGAGCGGCTGCC	732				
OY	760 GGCAATCAACCGTATGAGGTGTGAAATGCGCGGACAGATGGGCAATCAAGCACACCTG	819				
Dd	733 GGCAATCAACCGGAAACGCTGCTCGAGATGCGCGAATCGCTGCGACAGAGGGAATCTTC	792				

QY	CGACGCGCACACBACCCGGAACCTACGACCGCGACGATCTGTAATGGCGGGTACACACCGG	879
QY	820 CGACGCGCACACBACCCGGAACCTACGACCGCGACGATCTGTAATGGCGGGTACACACCGG <td>879</td>	879
Db	793 GCCGACATCACTGCTGCGCTCTCGACGCCGACGAATGCTCTCGCTGCTCACGACTGGG	852
QY	880 GCGCGGGTCACACCCGATCAACTCTGCTGATAGCGCGACGCGCGTGGGCAACGCGGACCGCGG	939
Db	853 GCGCGAGTGTGGCCATTTCGTACAGGTGTGACGCCAACCCCATCTCTCGGACGGGGTTCCCGGG	912
QY	940 CCACCTGACGGTGGCCATTCGCGGACCGGTTCTGGGCGCTGATGACACGACCGGCGCGCTG	999
Db	913 CCCATCAACCCAGTCGATCATCCGCTCTTACGTGGGAGCTGAATGTGAGAGCTCGTGTG	972
QY	1000 ATCGAAGCATCGAATACTGA	1020
Db	973 CTTAGCGCTGTGCACTACTGA	993

### RESULT 3

```

Sequence 7, Application US/10112706
Publication No. US20020192786A1
GENERAL INFORMATION:
APPLICANT: YAMADA, Yukio
APPLICANT: IWASAKI, Akira
APPLICANT: KITAZAKI, No. US20020192786A1,yuk1
APPLICANT: MATSUMOTO, Keiji
APPLICANT: IKENAKA, Yasuhito
APPLICANT: OGURA, Masahiro
APPLICANT: HASEGAWA, Junzo
TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
FILE REFERENCE: 1422-398P
CURRENT APPLICATION NUMBER: US/10/112,706
CURRENT FILING DATE: 2002-04-02
PRIOR APPLICATION NUMBER: 09/403,493
PRIOR FILING DATE: 1999-10-22
PRIOR APPLICATION NUMBER: JP 9-121732
PRIOR FILING DATE: 1997-04-23
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 7
LENGTH: 993
TYPE: DNA
ORGANISM: Arthrobacter sp.
US-10-112-706-7

```

Query Match	31.9%;	Score 325;	DB 9;	Length 993;
Best Local Similarity	58.2%;	Pred. No. 1.7e-74;		
Matches 571;	Conservative	0;	Mismatches 410;	Indels 0;
				Gaps 0;

QY	40	GGCGTGCAGACCGGGGGCCCAATTCGCGAGGACACCCGGGCGGGTCCGGTATTCAGTACAGC	99
Db	13	GCCGATACCTCCGAGATGCTCTACACGCACGACACCGGCTCTGACTACATCACTTATAC	72
QY	100	GACTACGAAGTGGAGACCTCCACGCCCGTTGCGCGCGCGCGTGCCTGGATCGAGGGCGAA	159
Db	73	GACTACGAAGTGGATCTCTCTTAACCGCGCTGCGGGAGTGGCGCAAGATCGAGGGTGC	132
QY	160	TACGTCCGCGCGGAGGAAGACGAAGATCTCATCTTCATGACACGGATTCGGTATTCGAT	219
Db	133	TTTCGTCCCGCCGTCGAGAGCGCCGGATCTCGATCTTCGATCAGGGTATCTCCACTTCGGAC	192
QY	220	CTGACCTACACCGTCGCGCATGTATGCGACGCGACATCTTTCGGCTCGGCCACCACTTG	279
Db	193	GTCACCTACACGCTCTTCCACGCTCGGAAGGAAATGCAATCTCCGCTCCACGACCCACATC	252
QY	280	GACCGTTCTCGACGCGGCGTCCAAAGTCTGGCGCTGGAAGCGCGGTATACAGACGACGAA	339
Db	253	GAACGCGCTCTTCTCCAACCGGGAGTGGATGCGGATATCCCTCCGTCACACGAGACGAA	312
QY	340	CTGGCCGAGATCACCAAGATGCTGTGTGATGTCCAGCTGGCGCAATCTTCGTGAAT	399
Db	313	GTGAAAGATATTCGCTCGAATCTCTGCGGAACCGCAATTCGTGAGGCTCTGTGTGC	372

OY	400	CTGACCGGCACCCGGGGATAGCAAAACGGAAAGGGGACGAAGAGACCTGTCCTCAAGCTCAC	4539
Dd	373	GTCGCGATTACCCGGGTTACGTCGACACTCCGGGGGACGACATACGAAGACACCCG	4322
OY	460	CATCAGGTGATCATCTACGCCATCCCGTACCTGTGGCGCTTCCCGCCGCCGACAGATC	5119
Dd	433	CCGGAGGTGTACATGTATGCCGTCCCATATCAGTGGATCTGTCCGCTTTGACCCGAATTCCG	4922
OY	520	TTCCGCACACCACGGGATCGTCGGCGGCATATGTCGGCGCGCGCGGCGCAACACCGTGCAC	5719
Dd	493	GACGGCGTGCAAGCGCATGTGTGCGCACAGAGCGTGGCGCGAACCOCGCGCACTGTGATGCAC	5522
OY	580	CCGACATCAAGAATCACTACAGTGGGGTGATTTCCACCGCAGCCAGCATTTTGAAGCCAAAGAC	6339
Dd	553	CCTCAGGTGCAAGAACTTCCATGTGGGGGGATCTGATCCGTGGGCTTCAAGAGACGACAGAC	6122
OY	640	CGTGGTCGGCGCACCGCGCATCTCTGCTGATCTGGACAACTGTCGTGGCGCGAAGATCCGGGC	6599
Dd	613	CGCGGGTTTCGAGGCGTCCCTCTTCTGCTGACGCGCATGAGACTGCTTGGCGAGGGCGTCCGGG	6727
OY	700	TTCAACGTGTGATCGTCAAGAGACGGCGAAGTGGGCTCTCCCGTCCCGGAAGCGGTTCCCG	7559
Dd	673	TTCAACGTGTGTCGTGATTAAGAGAGGGCGTGTGGCGAAGCCCGGGTCCAGCGGCCCTCCCC	7322
OY	760	GGCATCACCCGTTAAGACGGTGTTCGAACTGGCGGACCAAGATGGGCATCTGAAGCCACCCTG	8119
Dd	733	GGCATTTACGGGAAGAACCGTGTGCGAGATCCGCCAATCGCTCGGACACAGGCGCATTTTC	7922
OY	820	CGGACGTCACACGCGGTGAATCTACGACGCCGACAGATGATGATGGGGGTTCACACCCGCG	8719
Dd	793	GCCGACATCATCGTCGTGAACTGCTGCACGCCGACGAAAGTGTCTGGGCTACGACATCGCG	8519
OY	880	GGCGGGGTCACACCGATCAACTCGCTGGATGTCGAGAGCGCCGTGGGCAACGGGCAACCCGCT	9359
Dd	853	GGCGGAGTGTGGCCATTGTGCAGGGTGGAGGCAACCCCATCTCGGAGAGGGGTTCCGGGC	9122
OY	940	CCACTGACGGTGGCCATCCGGGACCGGTTCTTGGGCGCTGATGACAGACCGGGCCCGCTG	9959
Dd	913	CCCATCACCCAGTCATCATCTCGTGTTCACGTGGGAGGTGAATTCGAGAGCTCGTCTGTG	9727
OY	1000	ATCGAAAGATCGAATCTGA 1020	
Dd	973	CTTACCGCTGTGCATCTGA 993	

## RESULT 4

```

US-10-112-706-21
? Sequence 21, Application US/10112706
? Publication No. US20020192786A1
? GENERAL INFORMATION:
? APPLICANT: YAMADA, YUKIO
? APPLICANT: IMASAKI, Akira
? APPLICANT: KIZAKI, No. US20020192786A1yuk1
? APPLICANT: MATSUMOTO, Keiji
? APPLICANT: IKENAKA, Yasuhiro
? APPLICANT: OGURA, Masahiro
? APPLICANT: HASEGAWA, Junzo
? TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds
? FILE REFERENCE: 1422-398P
? CURRENT APPLICATION NUMBER: US/10/112,706
? CURRENT FILING DATE: 2002-04-02
? PRIOR APPLICATION NUMBER: 09/403,493
? PRIOR FILING DATE: 1999-10-22
? PRIOR APPLICATION NUMBER: JP 9-121732
? PRIOR FILING DATE: 1997-04-23
? NUMBER OF SEQ ID NOS: 26
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 21
? LENGTH: 1574
? TYPE: DNA
? ORGANISM: Arthrobacter sp.
US-10-112-706-21

```

Query Match 31.9%; Score 325; DB 9; Length 1574;  
 Best Local Similarity 58.2%; Pred. No. 1.7e-74;  
 Matches 571; Conservative 0; Mismatches 410; Indels 0; Gaps 0;

40 GCCGTCGAGCCGCGCATCCGCGAGAGACACCCGCGCGCGCTGGGATCCAGTACAGC 99  
 398 GCCGATACCTCCGAGATCGTCTACACGACAGACCCGCGCTGACTACATCACTATAGC 457  
 100 GACTACGAACTGAGACACCTCCAGCCGCTTCCGCGCGCGCTGGATGAGGCGAA 159  
 458 GACTACGAACTGATCTGCTAACCCGCTCGCGGAGGTGGGATGATGAGGAGTGA 517  
 160 TACCTGCGCGCGCGAGAGAGATCTCATCTTCGACACCGGATTCGGTCAATCCGAT 219  
 518 TTCTGCTCGCGCGCTCGAGGCGGATTCGATTCGATAGGATTAACCTCCACTCGAC 577  
 220 CTGACTACACCGCTCGCGCATGATGAGGAGGCAACATCTTCGCGCTCGCGACCACTG 279  
 578 GTGACCTACACGCTTCCTCACTGCTGAGAGGAAATGATTCATTCGCGCTCGACGACATC 637  
 280 GACCGGTTGCTCGAGGCGGCTCCAGCTGCGCTGAGCGCGGATACAGCAAGAGAA 339  
 638 GAGCGCTCTCTCTCCAGCGGAGTGCATGCGCATCCCTCCGCTCACACAGAGAGAA 697  
 340 CTGCGGAGATCCACAGAGATGGTGTGATGTCGACGCTGCGGCAATGCTTCGAT 399  
 698 GTGAGAGAGATGCGCTCGACTGCTGCGAAGACCGAATTCGCTGAGGCTTCGTC 757  
 400 CTGACCTGACCGCGGAGATACGGAAGCGAGAGGAGACCTGTCACAGCTCAC 459  
 758 GTGTCGATTCACCGCGGTTACAGCTGACCTCCGCGAGCGGCGACATACAGAAAGCAGC 817  
 460 CATGAGTGTACATCTAGCCATCCGCTACCTTGGGCTTCCCGCGCGCGAGAGATC 519  
 818 CCGAGGTGTACATGTATGCGCTCCATATCACTGATGCTGCGCTTTCGACCAATTCG 877  
 520 TTGCGGACCAACCGGATGCGCGCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579  
 878 GAGCGCTGACGCGCATGCTGCGACAGAGCGTCCGCGAAGCCCGCGAGCTGATCGAC 937  
 580 CCGACATCAAGAACTACAGTGGGTGATCTCAGCGACAGCTTTCGAGGCAAGAGAC 639  
 938 CCGAGGTGTACAGAACTCCAGTGGGAGATCTGATCCGCTGCGGTTCAAGAGAGCGAC 997  
 640 CCGTGTGCGGCGACCGGATCTGCTGACCTGCGACACTGCTGGCGCGAGTCCGGC 699  
 998 CCGGCTTCGAGGCTCCCTTCTGCTGACGCGGATGACCTGCTTCGAGGCTTCGGG 1057  
 700 TTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 759  
 1058 TTCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1117  
 760 GGCATCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 819  
 1118 GGCATCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1177  
 820 CCGGAGTACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 879  
 1178 GCGGATCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1237  
 880 GCGGAGTACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 939  
 1238 GCGGAGTACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1297  
 940 CCGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 999  
 1298 CCGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1357  
 1000 ATGGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
 1358 CTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1378

RESULT 5  
 US-10-112-706-5  
 ; Sequence 5, Application US/10112706  
 ; Publication No. US20020192786A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YAMADA, Yukio  
 ; APPLICANT: IMASAKI, Akira  
 ; APPLICANT: KIZAKI, No. US20020192786A1Yuk1  
 ; APPLICANT: MATSUMOTO, Keiji  
 ; APPLICANT: IKENAKA, Yasuhiko  
 ; APPLICANT: OGURA, Masahiro  
 ; APPLICANT: HASEGAWA, Junzo  
 ; TITLE OF INVENTION: Process for Producing Optically Active Amino Compounds  
 ; FILE REFERENCE: 1422-398P  
 ; CURRENT FILING DATE: 2002-04-02  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR FILING DATE: 1999-10-22  
 ; PRIOR APPLICATION NUMBER: JP 9-121732  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 5  
 ; LENGTH: 978  
 ; TYPE: DNA  
 ; ORGANISM: *Arthrobacter* sp.  
 US-10-112-706-5

Query Match 31.8%; Score 324.6; DB 9; Length 978;  
 Best Local Similarity 58.4%; Pred. No. 2.1e-74;  
 Matches 567; Conservative 0; Mismatches 404; Indels 0; Gaps 0;

50 CCGGCGCATCCGCGAGAGACACCCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 109  
 8 CCGAGATGTGTACACGACGACGACGACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 67  
 110 TGGACACTCCAGCCGCTTCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 169  
 68 TCGATCTGCTTACCCGCTGCGGAGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 127  
 170 CCGAAGAGCGAAGATCCATCTTCGACACCGGATTCGCTGCTGCTGCTGCTGCTGCTGCTG 229  
 128 CCGGAGGCGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187  
 230 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 289  
 188 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 247  
 230 TCGAGCGGCGCTCCAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 349  
 248 TCTCCAGCGGAGTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 307  
 350 TCACCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 409  
 308 TTGCGCTGGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 367  
 410 CCGGAGTACGGAAGCGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 469  
 368 CCGGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 427  
 470 ACATCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 529  
 428 ACATCTACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 487  
 530 CCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 589  
 488 ACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 547  
 590 AGAATACAGTGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 649  
 548 AGAATCTCACTGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 607  
 650 GCACCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 709



Db 1185 GGCTGACGTGCCCGACGCGCTTGCAGAGTGGCCGACCGGACCGGAACTCTCGCCG 1244  
 QY 524 GCACCAACCGCATGCTGCGCGCCATGTCGCCGCGCGCGGCGCAACACCGTCGACCGCA 583  
 Db 1245 GGTGACGCTGTTGCGCATGGCGGACCAACGCCCATGTGCTCTGACGGAAGGCCCG 1304  
 QY 584 CCATCAAGAACTACCAAGTGGGTGATCTCAGCGACGCCAGTTTGAAGCCAGAACGCGTG 643  
 Db 1305 CCCAGGGGCGGACGACCGCGGCATGATGAGAGACCCCGCTGACAGCGGCGCGCAC 1364  
 QY 644 GTGCGGACGACCGCATCTGCTGCTGACATGAGAACATGCGTGCGGAAGTCCGGGCTTCA 703  
 Db 1365 TGCCTTGTGCTGTCACCGCGCGCGCGGCGAGGCGCTCGCGGCCAGCGCGCGCTGC 1424  
 QY 704 ACCTGTGATGCTGTCAGAGACGCGCAAGTGGCGCTCCCGCGGAAAGCGTTCGCGGCA 763  
 Db 1425 ACAGAGCGGCTGCAAGCGGACCGCGGAGCTCGCGCGCGGACATCGCGCGCGCGTGC 1484  
 QY 764 TCACCGGTAAAGAGCTGTTGCAAGTGGCGGACCAAGATGGATGGAAGCCACTGCGCG 823  
 Db 1485 CCACCGGTACGCTTTCACGCAACGCGTGGTCTGCTCGCCGCGGACCGCGCGCGCTCC 1544  
 QY 824 ACCTCACCAAGCGGTGAACCTTACGACG 850  
 Db 1545 TCGACGGCTCGGCGCGCTCGCGCGCG 1571

## RESULT 8

US-09-988-384B-1

Sequence 1, Application US/09988384B

Publication No. US20030073824A1

GENERAL INFORMATION:

APPLICANT: Sherman, D. H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and plixromycin

FILE REFERENCE: 600.536US1

CURRENT APPLICATION NUMBER: US/09/988,384B

PRIOR APPLICATION NUMBER: PCT/US99/14398

PRIOR FILING DATE: 1999-06-25

PRIOR APPLICATION NUMBER: US 09/105,537

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 53

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-988-384B-1

Query Match 6.5%; Score 66.2; DB 9; Length 15872;  
 Best Local Similarity 44.4%; Pred. No. 5, 6e-08;  
 Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

QY 50 CCGGCGCATTCGCGAGAGACCCCGCGCTCGGATCCAGTACAGAGACTAGAAC 109  
 Db 765 CCGGCGCATTCGCGAGAGACCCCGCGCTCGGATCCAGTACAGAGACTAGAAC 824  
 QY 110 TGAACACCTTCAGCCGCTTTCGCGGCGGCGCTGCTGATGAGAGGCGAATACCTGCGG 169  
 Db 825 ACAGAGGAGGACACCCCGGCTCTACCGTGGCCAGAGAGGCGCGCGAGAGAGAGTCTCG 884  
 QY 170 CCGAG---AAGGAGATTCATCTTTCAGACACCGGATTCGCTCAATTCGATGAGCT 226  
 Db 885 GCGAGGAGTACCGGAGGCGCGCTGAGACCGCTCGCGCGCTGAGTACGTAACCTCAGG 944  
 QY 227 ACACGCTGCGCATGTATGAGCAGGCA---ACATCTTCGCGCTCGGCGAGCAACCTGAGAC 283  
 Db 945 GCAACCGAAGCCCGCTCGCGAGACCCCATGAGAGCGCGCGCTCGCGCTCGGCT 1004  
 QY 284 GGTGCTCGACGGGCGCTCAAGCTGCGCTGAGACCGCGGATACAGCAAGAGAACTGG 343

Db 1005 CGGCGCGCGCGCGGAGCAACCCCTGCTCGGCTCGGCGGCGCAAGAGAACTGCGGCGAC 1064  
 QY 344 CCGGATCACCAAGAACTGCTGCTGATGTCGACGCTGCGCGCATGCTTGTGTAATCTGA 403  
 Db 1065 TCGAAGGCGCGCGCGCATGCTGCGCTCTATCAAGAGAGCTCTCGGCTCGGCGCGCG 1124  
 QY 404 CCGTCAACCGGAGTACGAAAGAGCAAGAGGCGAGAGAGCTGTCTCAAGCTCACCCATC 463  
 Db 1125 GGAATCCCGGAGAGCTCACTTCCGTACGAGCCCAACCGGAGATCCGCTGACACCTTGC 1184  
 QY 464 AGGTGATCATCTAGCCATCCGTACCTGTGGGCTTTCCCGCGCGGAGCAAGATCTTGC 523  
 Db 1185 GGTGACGCTGCGCGGAGAGGCTCGGAGAGGCGCGGACCGGACCGCAACTCTCTGGCG 1244  
 QY 524 GCACACCGGAGATGTCGCGCGCATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1304  
 Db 1245 GGTGACGCTGCTTTCGCGATGGCGGCGGACCAACCGCGCGCTGCTCTCAGAGAGGCGCG 1364  
 QY 584 CCATCAAGAACTACCAAGTGGGTGATCTCACGCGACCGCAAGTTTGAAGCCAGGACCGTG 643  
 Db 1305 CCCAGGGGCGGACGACCGCGCGCATGATGAGAGAGAACCCCGCTGACAGCGGCGCGCAC 1364  
 QY 644 GTGCGGACGACCGGATCTGCTGACTGCGACAACTGCTGCGCGCGCGCGCGCTTCA 703  
 Db 1365 TGCCTTGTGCTGTCACCGCGCGCGCGGCGGAGGCGCTCGCGCGCGCGCGCGCTTC 1424  
 QY 704 ACCTGTGATGCTGTCAGAGACGCGCAAGTGGCGCTCCCGCGGAAAGCGTTCGCGGCA 763  
 Db 1425 ACAGAGCGGCTGCAAGCGGACCGGAGCTCGCGCGCGCGCGCGCGCGCTGCTGCTCA 1484  
 QY 764 TCACCGGTAAAGAGCTGTTGCAAGTGGCGGACCAAGATGGATGGAAGCCACTGCGCG 823  
 Db 1485 CCACCGGTACGCTTTCACGCAACGCGTGGTCTGCTCGCCGCGGACCGCGCGCGCTTC 1544  
 QY 824 ACCTCACCAAGCGGTGAACCTTACGACG 850  
 Db 1545 TCGACGGCTCGGCGCGCTCGCGCGCG 1571

## RESULT 9

US-09-861-289-1

Sequence 1, Application US/09861289

Patent No. US20020110897A1

GENERAL INFORMATION:

APPLICANT: Sherman, D. H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and plixromycin

FILE REFERENCE: 600.438US1

CURRENT APPLICATION NUMBER: US/09/861,289

PRIOR FILING DATE: 2001-05-18

PRIOR APPLICATION NUMBER: 09/105,537

PRIOR FILING DATE: 1998-06-26

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

US-09-861-289-1

Query Match 6.5%; Score 66.2; DB 10; Length 15872;  
 Best Local Similarity 44.4%; Pred. No. 5, 6e-08;  
 Matches 358; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

QY 50 CCGGCGCATTCGCGAGAGACCCCGCGCTCGGATCCAGTACAGAGACTAGAAC 109  
 Db 765 CCGGCGCATTCGCGAGAGACCCCGCGCTCGGATCCAGTACAGAGACTAGAAC 824  
 QY 110 TGAACACCTTCAGCCGCTTTCGCGGCGGCGCTGCTGATGAGAGGCGAATACCTGCGG 169  
 Db 825 ACAGAGGAGGACACCCCGGCTCTACCGTGGCCAGAGAGGCGCGCGAGAGAGTCTCG 884



```

QY 170 CCGAAG---AAGGAAGATTCATCTGACACCGGATTCGTCATTCGATCCGACT 226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 885 GCGAGGGGTACCGGAAGGGGGCCCTGGACCCGTCGCGCCGTCAGTACGCAATCCAG 944
QY 227 ACAACCGTCGCGATGTATGGACGSCA---ACATCTTCGGGCTCGCGGACCACTGGACC 283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 945 GCAACCGGAACCCCGTCGCGGAGACCCCATCGAGCCCGCCGCTCGCGGCTTCGGCT 1004
QY 284 GATTCGTCGACGGGGGTCGCAAGCTGCGCTGGACGCGGGTACAGCAAGAACGAACTG 343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 CCGGCGGCCCGCGCGAGCAACCCCTGCTGTCGGCTTCGGCCAAAGAACGTCGGGACC 1064
QY 344 CCGAATACCAAGAAAGTGGCTGTCATGTGCAAGCTGTCGCAAGCTTCACACCTG 403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1065 TCGAAGGCGCGCGGATGTCGCGCTCATCAAGACGCTCTCGCGCTCGCGGCGCGCG 1124
QY 404 CCGTCACCCCGGGATACGGAAGACCGCAAGGCGCAAGACCTGTCGCAAGCTACCCATC 463
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1125 GGAATCCCGGAGACCTCACTTCGTCAGCCCAACCGGACATCCGCTCGACACCTCG 1184
QY 464 AGGTGATCATCTAGCCATCCGCTACCTGAGGCTTCGCCCGCGCGGACAGATCTTG 523
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1185 GGGTCGACGTCGCGGAGGCTTCGCGGAGTGGCGCGCACCGGACCGCAACTGCTGCGCG 1244
QY 524 GCACACCGCGGATGTCGCGCGGCTGTCGCGCGCGCGCGCGCAACCGTCGACCGCA 583
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1245 GCGTCACCTGTCGCGGATGCGGCGGCGCAACGCGCCACGTCGTCAGAGGAAGCGCGCG 1304
QY 584 CCATCAAGAACTACGATGGGGTGATCTCAACCGCACGATTCGCAAGCCAAAGAACCGTG 643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1305 CCCAGGGCGCGAGACCCCGGATCGATGAGAGAACCCCGCGCAAGCGGGGCGCGCAC 1364
QY 644 GTGGCGCACCGGATCTGCTGATCTGAGCACTGCGAGCACTGCGGCGGAGTCCGCGCTTCA 703
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1365 TGCCCTTCGTCGTCACCGGCGCGCGCGGAGGCGGCGCGCGCGCGCGCGCGCGCTCG 1424
QY 704 ACGTGTCATGTCGAGAGGCGCAAGCTGGCTTCGCCGTCGCGGAAACGCTTCGCGGCA 763
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1425 ACGAGGCGCTGCAAGCGGACCGGAGCTCGCGCGCGCGCGCACTCGCGCTGCTGCTCA 1484
QY 764 TCACCGGTAGAGCGGTGTCGAACTGGCGGACGAGATGGGATGCGCAAGCCACCTGCGCG 823
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1485 CCACCGGTAGCGGTTCACGACGACCGGTCGTCCTCGCGCGGACCGCGCGCGCTTC 1544
QY 824 ACCTACCGCGGCTGCACTCTACGACG 850
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1545 TCGACGGCGTCGCGCGCTCGCGCGCG 1571

```

```

RESULT 10
US-09-918-740-57
; Sequence 57, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathways
; FILE REFERENCE: KAS-103X1
; CURRENT APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 57
; LENGTH: 6798
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Streptomyces sp Cl190 gene cluster containing mevalonate pathway
; OTHER INFORMATION: and IPP isomerase orfs

```

```

-US-09-918-740-57
Query Match 5.5%; Score 56.2; DB 9; Length 6798;
Best Local Similarity 45.9%; Pred. No. 2e-05;
Matches 228; Conservative 0; Mismatches 268; Indels 1; Gaps 1;

QY 464 AGGTATACATCTACCCATTCGCTGATCTGTGGGCTTCGCCCGCGCGGACAGATCTTCG 523
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1697 AAGCGACCTCGCTCTCAAGCGCGGAGCGGTCGCGCGGCTGACCTGACCGCGGCTGG 1756
QY 524 GCACACCGGATGTCGCGGCGCATGTGTCGCGCGCGCGCGCGGCAACCGTCGACCGCA 583
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1757 TCATGCGCGGTGTCAACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1816
QY 584 CCATCAAGAACTACGATGGGGTGTATCTACCGCACGCCAGTTTGAAGCCAAAGAACGCTG 643
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1817 CCGTCGACACCTCGCGCGGTGTACCGCGCGGTGGCGCGCATCGATGAAGACGACCTGACG 1876
QY 644 GTGGCGGACCGCGGATTCGCTGCTGACCTGGACAACTGGTGGCGGAGTCCGGGCTTCA 703
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1877 AGATGCGCTGCGGCTGCTGCGCGGCGGACCTGAGGCGGTGGCGGAGATCCGCGAGCGCA 1936
QY 704 ACGTGTCATCTCAAGAGCGCAAGCTGCGCTCCGCTTCGCGGAACGCTTTCGCGGCA 763
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1937 ACGCGCTGGGATGACAGCGCACCATGTGGCGCGCGCGCGCGCGCGGTGGGTACGTGCGC 1996
QY 764 TCACCGGTAGAGCGGTGTCGAACTGCGCGGACGACAGATGGGACATCGAACGACCTGCGCG 823
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1997 CCGCACAGGTCACCGCTGTCGACAGCTGCTCACTCGCGTCAAGAGACGATGTCCTGGCT 2056
QY 824 ACCTACCGCGGCTGCACTCTGACGCGCGGACGAGTTG-ATGGCGGTACACGCGCGGCG 882
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2057 ACGGACACATGAGACGCGGCTCCCAACGTGAGATGTCCTGTCGCGGCGGCGAGCGCGGCA 2116
QY 883 GGGGTCACACGATCACTGCTGATGCGGAGCGCGGTGGGCAAGCGGACCGCGTCCA 942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2117 GGGTGGCGGACGTCGATGACGCGCGCGCGCGCGCGGTGCGGTGAGTCTCTGCGCGGCGCG 2176
QY 943 CTGACGCTGCGCATCGG 959
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2177 GAGACGCTGCGCGCGCTG 2193

```

```

RESULT 11
US-09-918-740-63
; Sequence 63, Application US/09918740
; Publication No. US20030033626A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Frederick
; TITLE OF INVENTION: Manipulation of genes of the mevalonate and isoprenoid pathway
; FILE REFERENCE: KAS-103X1
; CURRENT APPLICATION NUMBER: US/09/918,740
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/221,703
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 63
; LENGTH: 8077
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Operon F containing A. thaliana, S. cerevisiae, and Streptomyces
; OTHER INFORMATION: DNA
US-09-918-740-63
Query Match 5.5%; Score 56.2; DB 9; Length 8077;
Best Local Similarity 45.9%; Pred. No. 2e-05;
Matches 228; Conservative 0; Mismatches 268; Indels 1; Gaps 1;

```

[illegible][illegible]

```

RESULT 14
US-09-974-300-492
: Sequence 492, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Berka, Randy M.
: TITLE OF INVENTION: Methods For Monitoring Multiple Genes
: FILE REFERENCE: 10085,500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: CURRENT FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: 09/680,558
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/279,526
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 492
: LENGTH: 810
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-492

```

[illegible]

Db	420	GTCCGGAATCGAGGCGCATATATGGCGGGGTGTGACCGGAAGCGCTCATGCTCATATGCATACAGGC	479
Qy	679	TGCGTGCCCGCAAGGTCCGGGCTTCACAGTGTGCATCGTCAAGGACCGCAACGGCGGCTCCG	738
Db	480		
Qy	739	CCGTCGCCGGA-----CGCGTTCCCGGGCATCACCGCTAAGACGGTGTTCGAACATGGCC	792
Db	540	CTTCGGGCTATATCGGAGCGCTTGAAAGGAATCACCCGGAAATGCATCATCAATAAAGG	599
Qy	793	GACCAAGATGGGCATCGAAGCCACCCTCGCGACGTCAACACCGGTGAATCTACAGAGCC	852
Db	600	CGAGAGCTCGGCTATGAAAGTGAAGAAGAGCGCGTTTACCGGCATGACGATACACAGCC	659
Qy	853	GACGATGATTGCGCGGTACCAACCGCGGGGGGTCAACCCGATCAACTGGCTGGATGGC	912
Db	660	GAGGAAAGTGTTTTAAACCGGAACCGCTGCAGAAATCTACCGCGGTCTGAAAAGTTGACGGC	719
Qy	913	GAGCGCCGTGGGCAACGGCGGACCGCGGTCCACTGACGGTGGCCATCCGGACCGGATCTTCG	972
Db	720	CGCAAGATCGGGGAGCGGCAAAACGGGAGCTCCACAAACCGGATGCTTGA AAAAGTTCCGC	779
Qy	973	GCGCTGATGAGAC	984
Db	780	GAGCCGCTCGTC	791

```

RESULT 15
US-10-145-415-98
; Sequence 98, Application US/10145415
; Publication No. US20030068788A1
GENERAL INFORMATION:
APPLICANT: Buckel, Thomas Gunter
APPLICANT: Hammer, Philip Eugene
APPLICANT: Hill, Dwight Steven
APPLICANT: Ligon, James Madison
APPLICANT: Molnar, Istvan
APPLICANT: Pachlatko, Johannes Paul
APPLICANT: Zirkle, Ross Eric
TITLE OR INVENTION: Methods and Compositions for Making Emamectin
FILE REFERENCE: STN-117 103846.312
CURRENT APPLICATION NUMBER: US/10/145,415
CURRENT FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: US 60/291,149
PRIOR FILING DATE: 2001-05-16
NUMBER OF SEQ. ID NOS: 105
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 98
LENGTH: 1266
TYPE: DNA
ORGANISM: Streptomyces
US-10-145-415-98

```

Query Match	5.0%;	Score 51;	DB 9;	Length 1266;
Best Local Similarity	43.9%;	Pred. No. 0.0004;		
Matches 311;	Conservative	0;	Mismatches 395;	Indels 3;
				Gaps 2;

QY	261	CCGCTGGGGGACACACCGTGGACCGGTGGTCGAGAGGGGGCGTCCAAAGTGGCGCTGGACGC	320
Db	207	CTGGATCCCCCAAGGACGACGATGCAACTGTGACCTGTGGCCACGCCGCGCTCCGCTCCACCC	266
OY	321	CGGGTACGACGAAGGACGAACTGGCCGAGATACCAAGAAAGTGGCTGTGATGTGCGACGT	380
Db	267	CGAGGGCAGAGAACGTCGCGCTCTGGCGAGAGGACCCCTGATCGCTACGACAACTGGTGTCT	326
OY	381	GGCGCAATCGTTCTGGAATCTGACCGCTACCCGGGGATAGGAGAAAGCGCAAGGGGCGAGAA	440
Db	327	GGCCACCGGCGCCGAACCGCGGCGCTGTGAATCCCGCGACCGGCTTGGCCCGCGTGC	386
OY	441	GGACCTGTCCAAGTCAACCACTATCAGGTGTATCATCTACGCATCCCGTACCTGTGGGCTTT	500
Db	387	CCACCTCGGCGCGCTCGGCCACGCGGAACGCGTCCGGGGGTCTCTGGCTCTCCCTCGGCGC	446

